

Portland State University
PDXScholar

Dissertations and Theses

Dissertations and Theses

Spring 5-18-2017

Preparing Historically Underserved Students for STEM Careers: the Role of an Inquiry-based High School Science Sequence Beginning with Physics

Jon P. Bridges
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Educational Leadership Commons](#), and the [Science and Mathematics Education Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Bridges, Jon P., "Preparing Historically Underserved Students for STEM Careers: the Role of an Inquiry-based High School Science Sequence Beginning with Physics" (2017). *Dissertations and Theses*. Paper 3626.

<https://doi.org/10.15760/etd.5518>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Preparing Historically Underserved Students for STEM Careers:
The Role of an Inquiry-based High School Science Sequence Beginning with Physics

by

Jon P. Bridges

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Education
in
Educational Leadership: Administration

Dissertation Committee:
Pat Burk, Chair
Bill Becker
Sybil Kelley
Deborah Peterson

Portland State University
2017

ABSTRACT

Improving the STEM readiness of students from historically underserved groups is a moral and economic imperative requiring greater attention and effort than has been shown to date. The current literature suggests a high school science sequence beginning with physics and centered on developing conceptual understanding, using inquiry labs and modeling to allow students to explore new ideas, and addressing and correcting student misconceptions can increase student interest in and preparation for STEM careers.

The purpose of this study was to determine if the science college readiness of historically underserved students can be improved by implementing an inquiry-based high school science sequence comprised of coursework in physics, chemistry, and biology for every student. The study used a retrospective cohort observational design to address the primary research question: are there differences between historically underserved students completing a Physics First science sequence and their peers completing a traditional science sequence in 1) science college-readiness test scores, 2) rates of science college-and career-readiness, and 3) interest in STEM? Small positive effects were found for all three outcomes for historically underserved students in the Physics First sequence.

ACKNOWLEDGMENTS

I could not have written this without the support of my oft-neglected family and friends. Thank you for giving me the support and space I needed on this journey.

The faculty and my fellow cohort members at Portland State have taught me a great deal. I hope it is reflected here. I am particularly indebted to Moti Hara for his guidance to design, conduct, and correctly interpret the results of a quality quantitative study to address the research question.

The encouragement and feedback from Deborah, Moti, Pat, Ryan, Sho, and Victor have shaped my thinking and helped me grow. Thank you for your time, patience, and unwavering dedication to equity.

TABLE OF CONTENTS

ABSTRACT.....	i
ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES.....	iv
LIST OF FIGURES.....	v
CHAPTER 1: STATEMENT OF THE PROBLEM.....	1
Background of the Problem	2
Research Problem	5
Significance	5
Research Methods and Question	8
Key Concepts	9
CHAPTER 2: A REVIEW OF RELATED LITERATURE.....	14
Theoretical Framework	14
Review of the Research Literature	17
Physics First.....	17
The opportunity to learn gap.....	19
Content coverage.....	22
Content exposure and emphasis.....	24
Instructional delivery.....	34
Review of the Methodological Literature	36
Summary of the Research Literature and Application to the Study	47
CHAPTER 3: METHODOLOGY.....	49
Research Methods	50
Participants	52
Procedures	53
Instruments and Measures	56
Role of the Researcher	65
Data Collection and Analysis	66
Conclusion	70
CHAPTER 4: RESULTS/ANALYSIS.....	72
Analysis of Data and Presentation of Results	73
Interpretation of Findings	93
Limitations of the Study	97
CHAPTER 5: DISCUSSION/CONCLUSION.....	100
Synthesis of Findings	101
The Larger Context	102
Implications	104
REFERENCES	112
APPENDIX A: CONTENT SPECIFICATIONS FOR THE ACT SCIENCE TEST.....	130
APPENDIX B: LINEAR REGRESSION MODEL.....	131
APPENDIX C: LOGISTIC REGRESSION MODELS.....	133

LIST OF TABLES

Table 1: Scientific and Engineering Practices	26
Table 2: Mechanistic and Organic Paradigms of Learning	28
Table 3: Demographics of 9 th Grade Students by Cohort	74
Table 4: Demographics of 9 th Grade Students Enrolled Three Years by Cohort	75
Table 5: Pearson Chi-Square Tests on Cohort Demographics by Enrollment Status	76
Table 6: 9 th Grade Demographics by Science Sequence	77
Table 7: Pearson Chi-Square Tests of Student Demographics by Science Sequence	77
Table 8: Z-Scores of Demographics of Students by Enrollment Status	78
Table 9: Mean Science Scores by Year	80
Table 10: College Ready in Science by Year	82
Table 11: Pearson Chi-Square Tests on 8th Grade Science College Readiness	82
Table 12: STEM Interest by Year	83
Table 13: Pearson Chi-Square Tests on 8 th Grade STEM Interest	84
Table 14: Science College Readiness Status in Grade 8 and 11 by Science Sequence	88
Table 15: STEM Interest in Grade 8 and 11 by Science Sequence	91

LIST OF FIGURES

Figure 1: Theoretical framework	16
Figure 2: Three of opportunity to learn (OTL)	20
Figure 3: Science college readiness testing by cohort	54
Figure 4: Research question outcome variables, methods, and predictors	70
Figure 5: Effect of Physics First on 11 th grade ACT science score	94
Figure 6: Effect of Physics First for economically disadvantaged students	95
Figure 7: Odds ratios produced by the binary logistic regressions	96
Figure 8: Grade 11 outcomes and effects of Physics First.....	102

CHAPTER 1 STATEMENT OF THE PROBLEM

Henry Levin (2009), who has conducted research over a span of more than 40 years on the economic benefits of investing in education, asserts “educational equity is a moral imperative for a society in which education is a crucial determinant of life’s chances” (p. 5). In *Rising above the Gathering Storm*, the National Academies (2007) note the rapid erosion in the U.S.’s competitiveness in science and technology—and thus the U.S. position as a global economic leader. At the same time, the U.S. population is in the midst of profound demographic change both in terms of the racial and economic composition of its citizens and the distribution of income among them. Public schools in America have historically had greater success educating middle-to-upper income and White students in math and science than historically underserved students (Kannapel & Clements, 2005). Thus, improving the STEM readiness of students from historically underserved groups is a moral and economic imperative requiring greater attention and effort than has been shown to date.

The current literature suggests a high school science sequence beginning with physics and centered on developing conceptual understanding, using inquiry labs and modeling to allow students to explore new ideas, and addressing and correcting student misconceptions can increase student interest in and preparation for STEM careers. This Physics First approach is grounded in constructivist learning principles and embeds aspects of culturally relevant pedagogy. The Next Generation Science Standards and the increased emphasis on preparing all students to be college- and career-ready in STEM provide a supportive policy environment for districts to adopt a Physics First approach.

However, successful implementation of a Physics First science sequence requires ensuring teachers have the content and pedagogical knowledge, self-efficacy, and the belief in the ability of all of their students to engage in rigorous science.

Background of the Problem

“Educational equity is a moral imperative for a society in which education is a crucial determinant of life’s chances” (Levin, 2009, p.5). Some scholars argue that “rather than ameliorating educational inequality” (Schmidt, Burroughs, Zioda, & Huong, 2015, p. 380), schools are exacerbating it. Others assert that schools have a “mixed and modest impact on the opportunity gap” (Putnam, 2015). While there are a number of both school and non-school factors that contribute to the achievement gap, I would argue that schools are morally obligated to increase efforts to reduce the gap in college and career readiness of Black and Latino students and their White and Asian peers, regardless of the sources.

The lifetime earnings of a college graduate are more than double those of a high school graduate for both genders and for all races (Levin, 2009), yet Black collegians are about half as likely and Hispanic/Latinos one-third as likely to earn a degree as their White peers (Deming & Dynarski, 2009). The median income among full-time workers with a bachelor’s degree in 2008 was \$55,700 compared to \$33,800 for high school graduates with no college degree and employed full-time (Baum, Ma, & Payea, 2010). Each new high school graduate generates over \$200,000 in economic benefit to society through increased tax revenue and savings in expenditures for health care, crime, and welfare (Levin, 2009). Even greater benefits accrue for high school graduates who

continue their education and pursue careers in science, technology, engineering and math (STEM) as well as bestowing “economic and other benefits on the nations and regions in which they live and work” (Zinth & Dounay, 2006, p.1). Unfortunately, historically underserved high school students are less likely to graduate from high school and are less likely to have sufficient opportunity to learn critical content, receive high quality instruction, and take high school courses that prepares them for college and pursuit of post-secondary study in STEM (ACT, 2015; Ottmar, Konold, Berry, Grissmer, & Cameron, 2013; Schiller, Schmidt, Muller, & Houang, 2010; Schmidt et al., 2015; Schwartz, Sadler, Sonnert, & Tai, 2009). This opportunity to learn gap has significant implications for individuals and for the country. Ornstein (2010) notes:

If the achievement gap in math and science had been closed between black and Hispanic students and white and Asian students by 1998 the Gross Domestic Product in 2008 would have been about \$400 to \$500 billion higher. If the gap between America’s low-income students and the remaining students had been similarly narrowed, GDP in 2008 would have been \$400 to \$670 billion higher. (p.426)

The U.S. position as a global leader may be abruptly lost without a greatly expanded commitment to achieving success in advanced education in STEM (National Academies, 2007). Only 15% of U.S. college graduates attain degrees in the natural sciences and engineering, compared to 50% in China (Freeman, 2008). It is estimated that the U.S. will need 1.75 million more engineers, a 20% increase, by the year 2010 (Gasbarra & Johnson, 2008). Demand for engineers is increasing at three times the rate of other professions (Gasbarra & Johnson, 2008), yet Blacks, Latinos, and Native Americans account for just 7% of science and engineering professionals in the United States, while constituting 25% of the U.S. population (Milloy, 2003).

At the same time, the U.S. is in the midst of profound demographic change both in terms of income distribution and in the racial and economic composition of its citizens. The poverty rate for young people under 18 in the U.S. rose from 16.7% to 21.8% in the years from 2002 to 2012 and more than a third of Hispanic/Latino and Black children live in poverty (Sparks, 2013). Poverty among children in America correlates to completing fewer years of schooling, working fewer hours and earning lower wages as adults, and a greater likelihood of reporting poor health and nutrition (Children's Defense Fund, 2012). The U.S. is projected to become a majority-minority nation for the first time in 2043 (U.S. Census Bureau, 2012). Sadly, public schools in America have historically had greater success educating middle-to-upper income and White students than poor students and students of color (Kannapel & Clements, 2005). The Children's Defense Fund (2012) warns that "more than three of four Black and Hispanic/Latino children, who will be a majority of our child population by 2019, are unable to read or compute at grade level in the fourth or eighth grade and will be unprepared to succeed in our increasingly competitive global economy" (p. 1). As the percentage of students living in low-income families has increased, the gap between the average reading and math skills of students from low- and high-income families (Duncan & Murnane, 2014; Reardon, 2013) and the gap in college graduation rates have increased substantially (Bailey & Dynarski, 2011). Nationally, the academic achievement gap as measured by the percentages of students meeting college readiness benchmarks in math and science is even more pronounced among Black and Hispanic/Latino students interested in STEM fields (ACT, 2015). There is little consensus on the primary cause of these educational disparities, but a body

of literature exists suggesting both in-school (e.g., schooling is geared toward the dominant White middle class culture, tracking practices, unequal distribution of quality teachers, low expectations for non-White and Asian students, normalization of failure) and out-of-school factors (e.g., family composition, poverty, student mobility, unequal per pupil expenditures, resegregation of neighborhoods) correlate with student achievement (Cowan Pitre, 2014; Jeynes, 2015).

Research Problem

While there are many factors contributing to disparate educational outcomes for low-income and Black and Hispanic/Latino students, the purpose of this study was to determine if the science college readiness of historically underserved students can be improved by implementing an inquiry-based high school science sequence comprised of coursework in physics, chemistry, and biology for every student.

Significance

The U.S. population will increasingly be comprised of citizens from racial or ethnic groups that are historically underrepresented in mathematics and science fields and who have historically underperformed on mathematics and science assessments (Zinth & Dounay, 2006). It is increasingly necessary that all workers are skilled in approaching math and science problems and solving problems (Center for Education Policy Analysis, 2008) as “the great majority of newly created jobs are the indirect or direct result of advancements in science and technology” (National Academies, 2010, p.18). Hence, the lack of preparation for and knowledge of STEM careers of historically underserved students both contributes to the threat of the global standing of the U.S. economy and

makes it less likely these students will benefit from the economic and intellectual rewards of a STEM career or high-skills, high-wage jobs in general.

U.S. students who report taking physics during high school are twice as likely to meet the ACT college readiness benchmark in science as other students (ACT, 2013c). Sadler and Tai (2007) find that high school courses in biology, chemistry, and physics prepare students for college courses in the same field. Unfortunately, only one in four Hispanic/Latino and Black students takes a physics course in high school; half the rate of their Asian peers (White & Tesfaye, 2011). One approach to closing this opportunity to learn gap in STEM for historically underserved students might be to increase the number of science credits required for high school graduation or require all students to complete biology, chemistry, and physics in order to graduate. However, evidence suggests that policies that increase science graduation requirements may not be effective alone for improving student outcomes (Buddin & Croft, 2014; Teitelbaum, 2003) or improving overall college enrollment rates or persistence (Montgomery, Allensworth, & Correa, 2010; Plunk, Tate, Bierut, & Gruzca, 2014). Further, increased graduation requirements in math and science increase the likelihood a student will drop out of high school (Plunk, Tate, Bierut, & Gruzca, 2014). Students who drop out do not benefit from increased science graduation requirements and the increased STEM readiness resulting from additional science coursework (Montgomery, Allensworth, & Correa, 2010).

An alternative to increasing science graduation requirements is to implement a Physics First sequence for all students. Physics First is a framework for a three-year core curriculum for high school science which inverts the traditional order in which science is

taught in high school so that physics is followed by chemistry and then biology (Pasero, 2001). The objective of the Physics First approach “is to build knowledge of science and the concurrent use of mathematics, following the hierarchical nature of science as it has unfolded over the past century” (Bardeen & Lederman, 1998, p.178). The beginning physics course in this inverted science sequence focuses on developing conceptual understanding rather than mathematical manipulation, uses inquiry labs to allow students to explore new ideas, and addresses and corrects student misconceptions (High School Committee of the American Association of Physics Teachers, 2009).

Despite better reflecting the evolving nature of science and how students learn, few schools begin their high school sequence with physics (Bardeen, & Lederman, 1998; Ewald, Hickman, Hickman, & Myers, 2005; Haber-Schaim, 1984). A 2005 survey of physics teachers found only 3% of public schools employ such a “Physics First” approach (High School Committee of the American Association of Physics Teachers, 2009). Pasero (2001) laments the lack of quantitative documentation of the outcomes of a Physics First approach and cites this absence as the most significant finding of his study of schools using Physics First. In the dozen years following Pasero’s call for more research on the outcomes of a Physics First approach, a handful of published studies suggest that Physics First increases students’ conceptual understanding of physics (Gaubatz, 2013; Liang, L., Fulmer, G., Majerich, D., Clevens, R., & Howanski, R., 2012; O’Brien & Thompson, 2009) and advanced science course-taking and achievement (Gaubatz, 2013; Goodman & Etkina, 2008; Liang et al., 2012; Livanis, 2006). While Gaubatz (2013) found no significant difference in mean ACT score gains

for students in a high school science sequence beginning with physics compared to their peers who took a traditional sequence, Dye, Cheatham, Rowell, Barlow, and Carlton (2013) found otherwise. There remains a dearth of research measuring the effects of a Physics First approach on the science college readiness of historically underserved students. Determining the effect of a Physics First approach on the science college readiness of students would inform the decisions of school and district leaders on how to better prepare their historically underserved students for post-secondary studies in STEM and STEM careers.

Research Methods and Question

This quantitative study utilized a retrospective cohort observational study design (Hoffmann & Lim, 2007; Mann, 2003) using extant data. The primary research question was: Are there differences between historically underserved students completing a Physics First science sequence and their peers completing a traditional science sequence in:

1. 11th grade ACT science scores;
2. College- and career-readiness as measured by the ACT science test; and
3. Interest in STEM as measured by the ACT interest inventory?

The study was situated in a suburban school district in the northwest United States with ten high schools. School enrollment varied from 200 to 2200 students with more than one-third of students qualifying for free or reduced price lunch and an equal number of students of color and White students. Students entering grade 9 in the fall of 2010 and 2011 served as the control group. The treatment group consisted of grade 9 students

entering in the fall of 2012 and the fall of 2013 (who were scheduled into an inquiry-based with modeling science sequence consisting of physics, chemistry, and biology). To reduce plausible rival explanations, statistical methods (i.e. multiple linear regression and binary logistic regression) were employed to adjust for initial differences between the two groups (if present) and reduce the effect of unwanted variables (Krathwohl, 2009). A number of confounding variables could not be controlled which weakened the internal integrity of the study, however. These variables included any differences between the two groups in science instruction received in the 8th grade between November and June and changes in teachers assigned to courses and their initial content and pedagogical knowledge. As with the study by Gaubatz (2013), findings from a study of a single district “should be tempered with the understanding that successful change within educational settings is context-dependent” (p. 25).

Key Concepts

For the purpose of this study, the following definitions are used:

Black is a term that refers to a person whose ethnicity is not Hispanic or Latino “having origins in any of the Black racial groups of Africa” (Humes, Jones, & Ramirez, 2011, p. 3). Blacks are people who indicated their race(s) as “‘Black, African Am., or Negro’ or reported entries such as African American, Kenyan, Nigerian, or Haitian” (Humes et al., 2011, p. 3).

College- and career-ready is a description of a high school student who possesses certain characteristics that are predictive of their success in college. It describes a student has achieved a score at or above a given college readiness benchmark on an EXPLORE,

PLAN, or ACT assessment (ACT, 2013d). College readiness benchmarks are scores on the ACT subject-area tests that represent the level of achievement required for a student to have a 50% chance of obtaining a B or higher or about a 75% chance of obtaining a C or higher in the corresponding credit-bearing first-year college courses of college algebra (mathematics) and biology (science) (ACT, 2013d). The college benchmarks for these tests are 18, 20, and 23, respectively. While these assessments are primarily content knowledge assessments, the college readiness benchmarks established for these tests predict a student's success in credit-bearing college coursework in the content area. Using a single assessment to predict college readiness is problematic, however (Maruyama, 2012). Other scholars define college and career readiness using different constructs and measures (Maruyama, 2012). David Conley's (2010) key dimensions of college readiness are college knowledge, academic behaviors, content knowledge, and key cognitive strategies. College knowledge, also referred to by Conley as "contextual skills and awareness," is defined as "the privileged information necessary to understand how college operates as a system and culture". Academic behaviors that generally relate to self-management is the dimension of college readiness that includes a "range of behaviors that reflects greater student self-awareness, self-monitoring, and self-control of a series of processes and behaviors necessary for academic success". Content knowledge is described as "overarching academic skills," which include reading and writing, and "core academic subjects knowledge and skills," encompassing English, mathematics, science, social studies, world languages, and the arts. Key cognitive strategies include problem formulation, interpretation, research skills, communication, and precision and

accuracy (Radcliffe & Bos, 2013). Spence (2007) defines college and career readiness as “the ability to read and write effectively and to think logically and symbolically, as taught in mathematics” (p. 42) while Burtnett (2010) writes that college and career readiness is “taking a core curriculum to prepare [students] for advanced career training or associate or bachelor's degrees — a ‘college-ready’ core of courses” (p. 42).

Economically disadvantaged students are students who are eligible for the National School Lunch Program.

Hispanic/Latino is a term that refers to “a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race” (Humes et al., 2011, p. 2).

Historically underserved students are Black, Hispanic/Latino, Native American, Pacific Islander, English language learners, or economically disadvantaged students. I use the term historically underserved to emphasize the patterns of STEM achievement for these student groups as opposed to inadvertently evaluating current educational programming for these students by using only the term underserved. Other scholars include different student groups in their definition of historically underserved and may substitute the word “underrepresented” for the word underserved. Hernandez, Schultz, Estrada, Woodcock, and Chance (2013) define underrepresented students in STEM as “women and African American and Latino students” (p. 89). Contreras (2011) adds Native American youth and subtracts women from his definition (p.505). For Shanahan, Pedretti, deCoito, and Baker (2011), students typically underrepresented in science are “English language learners (ELLs), girls, and students at low-achieving schools” (p. 131).

Native American is a term that refers to “a person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment” (Humes et al., 2011, p. 3).

Odds is “the probability of an event occurring divided by the probability of that event not occurring” (Fields, 2016, p. 880).

Odds ratio is the ratio of the odds of an event occurring in one group compared to the odds of the same event occurring in a second group (Durlak, 2009; Fields, 2016).

Opportunity to Learn (OTL) is the set of instructional activities provided to students to help them acquire the knowledge, skills, and abilities defined in a set of standards (Schmidt & McKnight, 2012). Tate (2008) deconstructs OTL into three aspects: 1) content exposure and coverage, 2) content emphasis, and 3) quality of instruction (Ottmar et al., 2013).

Pacific Islander is a term that refers to “a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands” (Humes et al., 2011, p. 3).

Physics First is a high school curricular framework for a three-year core curriculum for high school science which inverts the traditional order in which biology, chemistry, and physics are taught in high school so that physics is followed by chemistry and then biology (Pasero, 2001).

Scientific inquiry is a method of thinking that occurs when learners “construct explanations of phenomena in their world by generating questions, making predictions, marshaling evidence, building explanations, and integrating scientific concepts with real

world experience” (Marx, Blumenfeld, Krajcik, Fishman, Soloway, Geier, and Tali, 2004). Thadani, Cook, Griffis, Wise, and Blakely (2010) write that scientific inquiry is not only about the production of knowledge, but also critique. Moje (2007) defines scientific inquiry similar to Marx et al., but adds “communicating their findings to others” as an essential component (p. 11). Finally, Lederman, Lederman, and Antink (2013) also expand on the definition of Marks et al., writing “scientific inquiry extends beyond the mere development of process skills such as observing, inferring, classifying, predicting, measuring, questioning, interpreting and analyzing data [and] also refers to the combining of these processes with scientific knowledge, scientific reasoning and critical thinking to develop scientific knowledge” (p.142).

STEM is an acronym for a group of academic disciplines that are in the areas of science, technology, engineering, and mathematics (SEDTA, 2008). Other richer concepts of STEM include emphasizing an interdisciplinary approach to real-world lessons and/or contexts (Gerlach, 2012) or an integrated approach in which the four disciplines are integrated into a single course or courses (Brown, Brown, Reardon, & Merrill, 2011).

CHAPTER 2 A REVIEW OF RELATED LITERATURE

The purpose of this retrospective cohort observational study was to determine if the science college readiness of historically underserved students can be improved by implementing an inquiry-based high school science curriculum comprised of coursework in physics, chemistry, and biology for every student. The current literature indicates that student interest in and preparation for STEM careers may be increased by a Physics First approach centered on developing conceptual understanding, using inquiry labs and modeling to provide students opportunities to explore scientific phenomenon, and addressing and correcting student misconceptions.

Theoretical Framework

As the new century progresses, awareness has increased that high school diplomas too often leave young adults unprepared for success in college has increased (Conley, 2010; Darling-Hammond, 2010). In 2010, the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO) released the final Common Core State Standards (CCSS) which are “academic benchmarks intended to define the knowledge and skills that high school graduates will need to be successful in college and careers” (Center for Public Education, 2014, p. 16). To support schools and districts in providing all students with an internationally-benchmarked science education, teams from 26 states collaborated with a 41-member writing team and partners throughout the country to develop the Next Generation Science Standards (NGSS) which were released in April 2013 (NSTA, 2011). The NGSS “seek not only to provide students with a foundation of essential knowledge, but also to lead young people to apply their learning

through scientific inquiry and the engineering-design process to deepen understanding” (Robelen, 2013, p.1). The adoption of the NGSS, along with aligned curriculum and instructional materials, is a necessary but insufficient strategy for increasing the preparation of students for STEM careers (Bair & Bair, 2014). For students to attain the NGSS, educators must improve students’ opportunity to learn by focusing on “what and how well students are taught in classrooms” (Herman, 2007, p. 4). Darling-Hammond (2010) asserts that “unequal access to high-level courses and challenging curriculum explains much of the difference in achievement between minority students and White students” (p.52). Lee (2005) concurs by noting that when historically underserved students “are provided with equitable learning opportunities in school or in their communities, they demonstrate academic achievement, interest, and agency” (p. 438). After analyzing the 2012 Programme for International Student Assessment (PISA) results, Schmidt et al. (2015) conclude that “any serious effort to reduce educational inequalities must address unequal content coverage within schools” (p. 381). It logically follows that adoption of a science curriculum in which classes across schools are taught using common units of instruction aligned to standards would help address both across-school and in-school variation in content coverage. A science sequence with common standards and instruction would close the content gap, however I also assert that the instruction within such a sequence must be of high quality and culturally responsive. As Darling-Hammond reminds us:

Decades of research have shown that teachers who produce high levels of learning for initially lower- and higher-achieving students alike provide active learning opportunities involving student collaboration and many uses of oral and written language, connect to students’ prior knowledge and experiences, provide hands-

on learning opportunities, and engage students' higher order thought processes. (p.55)

Providing students with richer learning, inquiry-based instruction contributes to a socially-just pedagogy described by Moje (2007) and Thadani et al. (2010). I maintain that implementing a common science sequence for all students beginning with physics and grounded in inquiry and modeling (a Physics First approach) can close opportunity to learn (OTL) gaps experienced by traditionally underserved students in terms of both content and instructional quality while incorporating elements of a socially-just pedagogy. Further, successful implementation of a Physics First approach requires school and district actions to ensure teachers have the knowledge, self-efficacy, and beliefs to implement the new science curriculum sequence effectively (Asghar, Ellington,

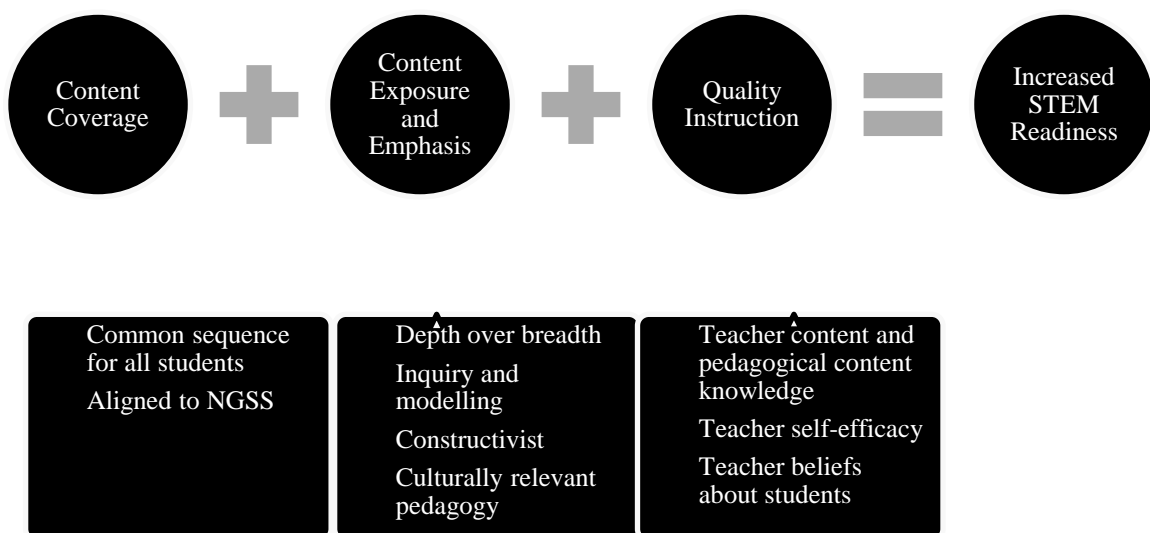


Figure 1. Theoretical framework.

Rice, Johnson, & Prime, 2012; Gibson & Brooks, 2012; Kessen & Henderson, 2010;

Lakshmanan, Heath, Perlmutter, & Elder, 2011; McGee, Wang, & Drew, 2013; National

High School Center, 2008; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). This theoretical framework for closing opportunity to learn gaps using a Physics First approach is shown in Figure 1.

Review of the Research Literature

I begin the review of the literature with a brief description of Physics First and its effects on improving student achievement in science. I then review the literature on opportunity to learn (OTL) gap and summarize the components of OTL. I conclude the review by relating research on aspects of Physics First to the three components of OTL described by Ottmar et al. (2013): content coverage, content exposure and emphasis, and instructional delivery.

Physics First. Physics First is a framework for a three-year core curriculum which inverts the traditional science sequence of biology, chemistry, and physics in U.S. high schools so that physics is taught first followed by chemistry and then biology (Pasero, 2001). The objective of the Physics First approach “is to build knowledge of science and the concurrent use of mathematics, following the hierarchical nature of science as it has unfolded over the past century” (Bardeen & Lederman, 1998, p.178). Understanding concepts of energy storage and transfer and electrostatic and nuclear forces in physics helps students master chemistry the following year. Likewise, students well-grounded in the basics of atoms and molecules developed in physics and chemistry will better understand DNA and polymers in biology. Uri Haber-Schaim (1984), an early proponent of Physics First, provides clear examples from science textbooks of

prerequisite knowledge from physics that are found in high school chemistry texts and the same for chemistry prerequisites found in biology textbooks.

A Physics First approach is not a mere reordering of the traditional high school science course sequence, however:

In a beginning course in physics, students explore their own notions about common, everyday phenomena, discuss their observations with peers, and draw conclusions to be tested. They begin to make predictions, practice data collection and graphing techniques, apply some mathematical skills to real situations, and start to make sense of observations.

(High School Committee of the American Association of Physics Teachers, 2009, pp. 6)

The beginning physics course in the inverted science sequence focuses on conceptual understanding rather than mathematical manipulation, uses inquiry labs to allow students to explore new ideas, and addresses and corrects student misconceptions (High School Committee of the American Association of Physics Teachers, 2009). The literature supporting the use of these instructional approaches to improve student learning in science are discussed further in a later section of this paper.

Despite better reflecting the evolving nature of science and how students learn, few schools begin their high school sequence with physics. A 2005 survey of physics teachers found only 3% of public schools employ a Physics First approach (High School Committee of the American Association of Physics Teachers, 2009). Pasero (2001) laments the lack of quantitative documentation of the outcomes of a Physics First approach and cites this absence as one of “the most significant findings” of his study of schools using Physics First (p. 13). In the dozen years following Pasero’s assessment, several studies suggest that Physics First increases students’ conceptual understanding of

physics (Gaubatz, 2013; Liang et al., 2012; O'Brien & Thompson, 2009) and increases advanced science course-taking and achievement (Gaubatz, 2013; Goodman & Etkina, 2008; Liang et al., 2012; Livanis, 2006). While Gaubatz (2013) identified no significant difference in mean ACT score gains for students in a high school science sequence beginning with physics compared to their peers who took a traditional sequence, Dye et al. (2013) found otherwise. Mean ACT science scores were higher for students in the Physics First sequence (effect size .14) and even higher for the Physics First with modeling (effect size .29). The increase in the percentage of students graduating college ready in science was roughly 20 percentage points higher for both the Physics First and Physics First with modeling compared to the traditional sequence (Dye et al., 2013). Unfortunately, none of these studies reported results for historically underserved students.

The opportunity to learn gap. The concept of opportunity to learn (OTL) is defined by researchers in a number of ways. Broadly speaking, opportunity to learn is the set of instructional activities provided to students to help them acquire the knowledge, skills, and abilities defined in a set of standards (Schmidt & McKnight, 2012). OTL originated in the work of the International Association of Educational Achievement during the 1960s to facilitate international comparisons of student achievement (McDonnell, 1995). Tate (2008) deconstructs OTL into three aspects: 1) content exposure and coverage, 2) content emphasis, and 3) quality of instruction (Ottmar et al., 2013). Stevens (1993) subdivides content exposure and content coverage into two distinct categories. Boscardin, Aguirre-Munoz, Stoker, Kim, Kim, and Lee (2005) deconstruct opportunity to learn into the facets of curriculum content, instructional strategies,

instructional resources (including teacher preparation), and assessment preparation developed by Herman, Klein, and Abedi (2000). Content coverage refers to the extent to which students are exposed to the core concepts identified for their grade or class. Content exposure encompasses both time devoted to instruction and depth of teaching. Content emphasis refers to both which topics receive emphasis and the balance in instruction of lower-order and higher-order skills. Finally, instructional delivery examines factors such as coherence, the quality of interactions between students and teachers, and the pace of instruction. (McDonnell, 1993; Stevens, 1993; Wang, 1998).



Figure 2. Three of opportunity to learn (OTL).

Over the past half century, numerous studies have demonstrated the link between OTL illustrated in Figure 2 and student achievement in mathematics and science (Schmidt et al., 2015). In a study of the relationship between OTL and student achievement on high school end-of-course exams in Algebra and English, Boscardin et al. (2005) found that content coverage was positively correlated with student performance. Ottmar et al. (2013) found an association between content exposure and achievement of 5th grade students in mathematics. The “pedagogy of poverty” experienced by student in many urban classrooms described by Haberman (1991)

encompasses low quality instruction with an emphasis on lower-order thinking and low-level content, often delivered by less prepared teachers (Berry, Ellis, & Hughes, 2013; Stevens, 1993). For historically underserved students and Black students in particular, this opportunity to learn gap translates directly to an achievement gap (Berry et al., 2013; Tate, 2008). Other researchers, however, point out other gaps outside the school setting, including residential segregation, unequal access to health care, differences in community assets, family structure and parental support, unequal access to community resources that influence the low achievement of historically underserved students (Ladson-Billings, 2006; Milner, 2012; Putnam, 2015). While these external influences undoubtedly influence student achievement, they cannot be directly addressed by schools. Conscious efforts to close opportunity to learn gaps are within the purview of schools. As Rotherham and Willingham (2009) assert:

Today we cannot afford a system in which receiving a high-quality education is akin to a game of bingo. If we are to have a more equitable and effective public education system, skills that have been the province of the few must become universal (p. 16).

As demonstrated by Schmidt and McKnight (2012), this game of bingo and its prize of access to high quality opportunities to learn occurs at all levels of education (between communities, between schools, and between classrooms) with the greatest source of variation in opportunity to learn occurring between classrooms. Banks, Cookson, Gay, Hawley, Irvine, Nieto, Schofield, and Stephan (2001) argue schools must “ensure that all students have equitable opportunities to learn and to meet high standards” (p. 198).

Content coverage. The creation and adoption of common standards by states provide an impetus for reducing the variation across states, schools, and classrooms by defining common content coverage for all students and thereby reduce inequality in opportunities to learn (Schmidt & Burroughs, 2013). However, the number of science courses students have to choose from undermines this attempt at closing the opportunity gap. As documented by Schmidt and McKnight (2012) in their analysis of course sequences in 16 districts participating in the Third International Math and Science Study replication (TIMSS-R), the number of science courses offered in high school ranged from 7 to 55. This variety in science courses results in many possible sequences or tracks students may experience in high school, “leading to very different learning experiences with science content” (Schmidt & McKnight, 2012). Students in different sequences or tracks have access to different types of content knowledge and experience different types of classroom instruction (Abedi & Herman, 2010; Callahan, 2005; Oakes, 1990). Simply put, tracking promotes inequality:

The achievement gap between students in high-level classes and those in low-level classes grows over time. A major cause for increasing inequality is that the pace, complexity, and challenge of classroom instruction are higher in high-track classes than elsewhere (Gamoran, Porter, Smithson, & White, 1997, pp.325-326).

English Language Learners, Black, Latino, and economically disadvantaged students are underrepresented in high track classes, even after controlling for prior achievement (Burris & Wellner, 2005; Callahan, 2005; Oakes, 1990; Oakes & Wells, 1998; Zuniga, Olson, & Winter, 2005). Assigning students to a common high school science sequence would close OTL gaps in content coverage. However, proponents of ability tracking argue that tracking improves learning for all students by reducing the extreme variation in

student abilities within individual classrooms (Zuniga et al., 2005). In their analysis of 10th graders in the National Educational Longitudinal Study of 1998, Argys, Rees, and Brewer (1998) found that low-ability students improved on achievement tests when assigned to heterogeneous math classes while average and high-ability students lost ground. Another study from the same data set found that all students learned more in tracked algebra classes than in heterogeneously grouped classes (Loveless, 1999). In contrast, Oakes (1990) cites numerous studies demonstrating that high-ability students do as well in mixed-ability classes as in tracked classes.

It should be noted here that de-tracking or closing gaps in content coverage alone may not fully close the opportunity to learn gap for historically underrepresented students. Based on their ethnographic study of a low-track Earth science classroom in a southern high school, Gilbert and Yerrick (2001) warn that detracking schools will not bring about positive results as long as school structures that rely on punitive means to maintain student compliance with school rules rather than resolve conflicts are not addressed.

Research on high school course-taking patterns has shown that enrollment in advanced-level science and math courses is related to college aspirations, college attendance and degree attainment (Tyson, Lee, Borman, & Hanson, 2007). In their study of course-taking patterns among Florida high school graduates, Tyson et al. (2007) found that Black and Hispanic/Latino students who complete advanced science and math coursework are at least as likely to obtain a STEM degree in college as their White counterparts. I argue that a Physics First approach in which all students complete

physics, chemistry and biology during their first three years of high school reduces content variation and opens the door for more students, particularly historically underserved students, to take advanced coursework in science. This approach answers the call of Schmidt & McKnight (2012) for “a sensible, limited number of tracks (sequences) that are less arbitrary and that reflect 21st-century economic realities” (p. 122). I further assert that a Physics First approach for all students supports student success in science coursework in college. Students earn higher grades in college biology, chemistry, and physics courses if they have taken the same subject in high school compared to students who have not had the corresponding course in high school (Sadler & Tai, 2007).

Content exposure and emphasis. Compared to other countries, the K-12 math and science curriculum in the United States has been criticized for having too many topics that are taught at a superficial level (Krajcik & Merritt, 2012; Robelen, 2010; Schmidt, Wang, & McKnight, 2005). *A Framework for K-12 Science Education* (National Research Council, 2012), on which the NGSS are based, is an attempt to bring coherence and depth to the U.S. science curriculum. The second dimension of OTL, content emphasis and exposure, is concerned with depth of teaching, topic coverage, and the balance in instruction of lower-order and higher-order skills. Like the debate over tracking, aspects of this dimension of OTL –depth versus breadth in the curriculum and how students best learn -- have been debated for decades. In this section I examine the literature on depth versus breadth, scientific inquiry and modeling, the learning theory of Physics First, constructivism, and culturally relevant pedagogy.

Depth versus breadth. Time spent on topics and the cognitive demand of tasks are key components of quality curriculum and opportunity to learn (Schiller et al., 2010). Students in lower-level math tracks “receive lesser amounts of cognitively challenging instructional material in their courses than those in regular or advanced tracks” (Schiller et al., 2010, p.428). Compared to their international peers, U.S. students and their teachers work with math and science textbooks that have more topics, a wider variety of topics across textbooks, and have more content breaks (Schmidt & McKnight, 2012). Textbooks determine the content for 75% to 90% of classroom instruction across the United States (Boone, 2006; Finn & Ravitch, 2001). The emphasis on breadth over depth in U.S. textbooks makes learning science and math more difficult for students in the U.S. (Schmidt & McKnight, 2012). Further, students who race through more content in textbook-centered courses do less well in college coursework than students who report spending more time on fewer topics (Sadler & Tai, 2007; Schwartz et al., 2008; Tai, Sadler, & Mintzes, 2006). Sadler and Tai (2007) also found that students whose high school science teachers emphasize conceptual understanding earn higher grades in college science coursework. Scientific inquiry and modeling as methods for increasing the conceptual understanding of students are discussed in more detail below.

Scientific inquiry and modeling. Scientific inquiry is a method of thinking that occurs when learners “construct explanations of phenomena in their world by generating questions, making predictions, marshaling evidence, building explanations, and integrating scientific concepts with real world experience” (Marx et al., 2004). Scientific inquiry by students reflects the same type of behavior that real-life scientists use,

although “not on the same scale” (Lederman, 1998, p.10). Inquiry pedagogy is prominent in the national standards documents promulgated during the last twenty years. The National Research Council’s *A Framework for K-12 Science Education*, which serves as the foundation for the Next Generation Science Standards, describes scientific inquiry in the Scientific and Engineering Practices dimension of the *Framework* listed in Table 1.

The inquiry-oriented instruction outlined in *A Framework for K-12 Science Education* positive impacts student achievement (Kanter & Konstantopoulos, 2010; Schroeder, Scott, Tolson, Huang, & Lee, 2007). The meta-analysis conducted by Minner, Levy, and Century (2009) also finds a clear, positive trend favoring inquiry-based instructional practices, particularly pedagogy that emphasizes student active thinking and drawing conclusions from data. As reported by Geier et al. (2008), the implementation of standards-based, inquiry science units also leads to standardized achievement test gains for historically underserved urban students. Thus, inquiry instruction is a viable strategy for closing the achievement gap in science (Johnson, 2009; Marshall & Alston, 2014).

Table 1: Scientific and Engineering Practices

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information

A Framework for K-12 Science Education (National Research Council, 2012, p. 42)

Modeling instruction emphasizes students constructing and applying conceptual models of physical phenomenon in order to learn science. Jackson, Dukerich, and Hestenes (2008) describe the essence of modeling instruction and provide strong evidence of effectiveness of the modeling approach. The modeling cycle consists of two stages: model development and model deployment. In the model development stage, the teacher sets the stage with a class demonstration and discussion related to a key question. Students then work in small groups to plan and conduct experiments to answer or clarify the question. In the second stage, model deployment, students deploy their newly-formed model to a new situation in order to refine and deepen their understanding of the concept. Jackson et al. assert that modeling instruction corrects many of the deficiencies of traditional science instruction, including fragmented knowledge, student passivity, and the persistence of student misconceptions. Data from the Force Concept Inventory (FCI) “the most widely used and influential instrument for assessing the effectiveness of introductory physics instruction” provide evidence supporting the effectiveness of modeling instruction compared to traditional instruction (Jackson et al., 2008, p. 15). The results from over 30,000 students taking the FCI as both a pre- and post-test found that students in modeling classes demonstrated a gain of more than double the learning of students in traditional instruction (Jackson et al., 2008).

The hands-on, minds-on nature of inquiry and modeling can also increase the interest of students in science. Examining the experience of students in a chemistry class in an alternative high school, Peterson-Beeton (2007) reports that Latino students in the school lose interest in science due in part to the lack of hands-on activities. Review of the

literature suggests that inquiry and modelling are also compatible with both brain science and constructivist learning as summarized in the next two sections.

Learning theory of Physics First. In proposing an inverted science sequence beginning with physics, Lederman (1998) draws upon the neuro and cognitive sciences to reject a “mechanistic” paradigm of learning in favor of an “organic” one as outlined in Table 2 (p. 8). According to Lederman (1998), this organic approach to learning requires that science teachers create conditions for learning that enable students to:

- process many different kinds of information simultaneously;
- understand information when it is embedded in messy yet relevant, authentic, novel, challenging and information-rich contexts;
- construct meaning through connections and pattern formulation;
- organize and associate new information with their existing knowledge;
- collaborate with peers and adults in challenging (but not threatening) endeavors; and
- actively and continuously engage in the practice of their new learning by constantly revisiting it at increasingly higher levels of complexity over extended periods of time (p. 8-9)

Table 2: Mechanistic and Organic Paradigms of Learning

Mechanistic Paradigm of Learning	Organic Paradigm of Learning
The brain as serial computer	The brain acts as a parallel processor able to process many different kinds of information simultaneously.
Learning as information accumulation	Learning is an internally and socially mediated process of constructing meaning from patterns created through multiple representations of knowledge
The mind as a <i>tabula rasa</i>	The mind is a dynamic, self-organizing “plastic” neural network that learns best when the context of learning is embedded in the entire physiology—including the body and the emotions.

An inverted science sequence beginning with physics creates these conditions by reflecting the nature of modern science. Modern biology courses emphasize genetics, molecular methods, and biochemistry (O'Brien & Thompson, 2009). Modern chemistry emphasizes atomic structure (O'Brien & Thompson, 2009). Understanding modern chemistry requires a solid grounding in physics while a good understanding of modern biology requires fundamental understanding of principles and concepts of both chemistry and physics (Haber-Schaim, 1984; Liang et al., 2012; Mervis, 1998). Physics, "the most concrete of sciences," provides a platform for students to understand "the unobservable interactions between atoms and molecules" fundamental in chemistry and biology (Hill, 2013, p. 38).

Constructivism. The conditions of learning in Lederman's organic approach are consistent with constructivist learning principles: assimilation of knowledge into current knowledge structures/schema, collaborating with peers and more knowledgeable others during learning, challenging tasks appropriate for the learner's Zone of Proximal Development, and extending learning at higher levels of complexity over time (Lutz & Huitt, 2004).

Cakir (2008) posits that the growth in the use of constructivist pedagogy may be ascribed to the appeal of aspects of the constructivist learning theories of Piaget, Ausubel, and Vygotsky outlined above; "namely, the importance of ascertaining prior knowledge, or existing cognitive frameworks, as well as the use of dissonant events (relevant information) to drive conceptual change" (p. 196). Based on his review of the literature, Cakir argues that science teachers would be more effective if "they understood the

barriers to conceptual learning (particularly the strong hold of prior misconceptions and the resistance to conventional instruction) and if they become familiar with the education research and strategies dealing with these misconceptions” (p. 202). A study of 181 middle school science teachers conducted by Sadler, Sonnert, Coyle, Cook-Smith, and Miller (2013) supports Cakir’s assertion. Sadler et al. found that teachers’ understanding of their students’ most common misconceptions was correlated with increased gains in the learning of their students. In addition to providing learning experiences which directly confront student misconceptions, Cakir suggests teachers: 1) recognize that science concepts are learned over time (through integration into student’s existing schemata); 2) understand effective science lessons are a social process in which the teacher and peers play a crucial role; 3) recognize the role of language in conceptual development; and 4) understand that learning science involves students being initiated into the ideas and procedures of the scientific community (enculturation) as well as making these ideas meaningful at the individual level of the student.

While behaviorists would decry Lederman’s rejection of the mind as *tabula rasa*, his organic learning principles align with Haberman’s concept of “good teaching.” Haberman (1991) describes both the “pedagogy of poverty” experienced by students in high poverty, high minority urban schools and its alternative, good teaching. The pedagogy of poverty is characterized by teacher-directed classrooms in which the core functions of a teacher are giving information, directions, assignments, and homework; monitoring student behavior and seatwork; and marking papers and giving grades. The result of the pedagogy of poverty is “nonthinking, underdeveloped, unemployable”

citizens representing a “personal and societal tragedy” (Haberman, 1999, p. 294). The alternative to this pedagogy of poverty, good teaching, is comprised of student involvement in vital issues; explanations of human differences; instruction organized around big ideas; students applying ideals of fairness and equity; active student participation in planning and lessons; real-life experiences; students working in heterogeneous groups; and opportunities for students to polish work, critique big ideas, and reflect on the personal meaning of their learning. Haberman’s conception of good teaching is echoed over a decade later in the guiding principles of *A Framework for K-12 Science Education*, a foundation for the Next Generation Science Standards:

These principles include young children’s capacity to learn science, a focus on core ideas, the development of true understanding over time, the consideration both of knowledge and practice, the linkage of science education to students’ interests and experiences, and the promotion of equity. (National Research Council, 2012, p. 24)

Thadani et al. (2009) examined the role that curriculum-based inquiry interventions in science might play in addressing Haberman’s pedagogy of poverty. By providing students with richer learning, inquiry-based instruction contributes to a socially just pedagogy described by Moje (2007). Thadani et al. also describe how science inquiry contributes to a social justice pedagogy:

Inquiry apprentices children into scientific practice by teaching them to generate questions and reason from and about evidence. . . And by positioning children as either producers or critics of scientific knowledge, inquiry-based learning disrupts traditional teacher-student roles. Students are required to take responsibility (albeit to varying degrees, in different inquiry projects) for their work. Their ideas (rather than teachers’ ideas or the ideas of some other scientific authority) become the central subject of discussion. Moreover, to the extent that inquiry-based instruction requires students to generate arguments and critique their own and each other’s ideas, it again disrupts the ‘teacher in charge’ model of instruction (i.e., the social context) that is emblematic of the pedagogy of poverty. (pp. 23)

Evaluating the effects of an inquiry-based intervention in three classrooms, Thadani et al. found teaching in intervention classrooms was more inquiry-based and less didactic than in control classrooms with differences in control/intervention teaching greatest at the two higher-need schools. Intervention students at these two schools had greater gains in content learning (effect size of .14 and .22) than control students in each school. The authors posit their findings support the “potential power of inquiry-based teaching for challenging the pedagogy of poverty” (Thadani et al., 2009, p. 35). At the same time, Thadani et al. warn that access to science inquiry pedagogy curricula alone will not remedy the achievement gap in STEM: “teachers and students who are most entrenched in it [the pedagogy of poverty] are likely to have more difficulty using such interventions faithfully because inquiry-based practices run so counter to their prior experiences and beliefs” (p. 35).

Critical theory. Critical theorists would argue that even if the pedagogical inequities described earlier are addressed through the faithful use of inquiry-based science teaching practices, historically underserved students will not achieve at the same level as their peers in the dominant culture. Lee and Buxton (2011) document three theoretical perspectives that have been applied to “the challenge of providing engaging and equitable science opportunities” for historically underserved students: (a) a cognitively based perspective, (b) a cross-cultural perspective, and (c) a sociopolitical perspective (p. 278). I assert that a Physics First approach with modeling is well aligned with a cognitive perspective of engaging historically underserved students in science through the use of deep questions and vigorous argumentation. Students learn to use

language, to think, and to act as members of a scientific community. A Physics First approach also provides all students with equal access to high quality, inquiry driven instruction. However, I also maintain that the cross-cultural and socio-political perspectives described by Lee and Buxton (2011) and advocated by Banks et al. (2001) are largely unaddressed.

Many educators assert that the incorporation of culturally relevant pedagogy into a Physics First approach would address these cross-cultural and socio-political perspectives. For Ladson-Billings (1995), culturally relevant pedagogy rests on three propositions: (a) teachers use high quality instruction to develop academic skills so that students experience academic success; (b) teachers use students' culture as a vehicle for learning while students maintain their cultural identity; and (c) teachers provide students with opportunities to critically analyze society so that students develop a critical consciousness through which they challenge the status quo of the current social order. As I argued earlier, a Physics First science sequence for all students incorporating inquiry and modeling develops the academic skills of all students through active learning and the construction of new knowledge using students' prior experiences and misconception. "Goals of equity and social justice lie at the core" of this Physics First approach to science (Bardeen & Lederman 1998). Engaging with peers in argumentation from evidence also draws upon the culture of historically underrepresented students (Kanter & Konstantopolous, 2010; Lee & Buxton, 2011). Reframing science as constructing meaning instead of information acquisition is also culturally congruent and redistributes authority within the classroom (Kanter & Konstantopolous, 2010; Patchen & Cox-

Petersen, 2008). Nonetheless, I concede that the development of the critical consciousness of students remains divorced from the Physics First approach described here.

Instructional delivery. For the final domain of OTL, instructional delivery, I examine the literature related to teacher content and pedagogical content knowledge, the beliefs of teachers in their own teaching ability and their ability to impact student learning (self-efficacy), and teachers' beliefs about the ability of their students to engage in the science as envisioned by the NGSS. Implementing Physics First requires science teachers to have strong content knowledge and possess the pedagogical knowledge to implement modeling and inquiry effectively with all students (Asghar et al., 2012; Gibson & Brooks, 2012; Kesson & Henderson, 2010; McGee et al., 2013; National High School Center, 2011; Penuel et al., 2007). Converting to a Physics First approach will also require disruption of the teaching assignments of some science teachers in addition to the time and expense of professional development (Mervis, 1998; Popkin, 2009; Taylor et al., 2005). Based on the literature, I argue that professional learning opportunities addressing teacher knowledge and beliefs increases the likelihood of successful implementation of Physics First and thus improvement in the science college readiness of historically underserved students.

Developing teacher content and pedagogical content knowledge. Implementing an inverted science sequence grounded in inquiry and modeling requires professional development for teachers who may or may not have strong content backgrounds due to reassignment and/or incomplete understanding of how students learn science (National

Academies of Sciences, Engineering, and Medicine, 2015). Further, “many teachers learned to teach using a model of teaching and learning that focuses heavily on memorizing facts, without also emphasizing deeper understanding of subject knowledge” (Garet, Porter, Desimone, Birman, & Yoon, 2001, p. 916). Wallace and Kang (2004) assert that “teachers’ understanding of the nature of science may create barriers to implementing inquiry-based instruction” (p. 940). High quality professional development that enhances teachers’ understanding of the nature of science, strengthens the content knowledge of science teachers and the pedagogical knowledge of how to teach the new physics course using inquiry and modeling methods is required (Asghar et al., 2012; Gibson & Brooks, 2012; Kesson & Henderson, 2010; McGee et al., 2013; National High School Center, 2008; Penuel et al., 2007). Studies indicate that professional development opportunities for increasing content knowledge and pedagogical knowledge are motivating for teachers (Anderson, 2008; Fields, Levy, Karelitz, Martinez-Gudapakkam, & Jablonski, 2012).

Developing teacher’s sense of self-efficacy. In addition to strong science content and pedagogical knowledge, teachers must also believe in their own ability to change student achievement outcomes for their historically underserved students.

For teachers to learn a new set of competences to help them leave fewer children behind in their classrooms, they may have to endure a temporary loss of confidence as they face the gap between the demands for performance and their current practice. To tell a teacher that she has to begin measuring her success by how well she raises test scores or teaches ‘unteachable’ students may challenge a great deal about what she was taught about her job. (Heifetz & Linsky, 2004, p. 35)

“Change involves learning to do something new” (Fullan, 1994, p. 2843). Asking teachers to adopt new teaching techniques may engender feelings in teachers of being de-

skilled (Altrichter, 2005). Gibson and Brooks (2012) note the tension between the genuine desire of teachers to improve practice and the need to maintain feelings of overall competence and self-efficacy. Self-efficacy refers to the belief in one's ability to successfully perform a task while efficacy refers to the belief in the ability to impact student learning (Lakshmanan et al., 2011). Lakshmanan et al. observe that both high teacher self-efficacy and efficacy have been linked to increased student achievement. Teacher self-efficacy and efficacy can be increased through professional learning (Lakshmanan et al., 2011). Embedded professional learning opportunities for teachers can support their adoption of new teaching practices (Camburn, 2010; Hunzicker, 2012). Both Bair and Bair (2014) and Wallace and Kang (2013) argue that sustained and supported professional development is integral for enhancing teacher skills in implementing inquiry-based science. Employing student centered approaches including inquiry and modeling can increase teacher self-efficacy (Hunzicker, 2012).

Teacher beliefs about their students' abilities to learn science. When historically underserved students “are provided with equitable learning opportunities in school or in their communities, they demonstrate academic achievement, interest, and agency” (Lee, 2005, p. 438). However, as noted by Anderson in Larkin, Seyforth, and Lasky (2014), “many teachers see a tension between providing a strong education for the able and willing students and at the same time providing for the uninterested or less able students” (p. 828). Further, Wallace and Kang (2004) assert that “teacher beliefs about the limitations of their students in terms of ‘ability’ or ‘maturity’ can be an obstacle to more student-centered approaches to instruction” such as scientific inquiry (p. 940). To

remove this obstacle, I assert that professional development must explicitly include the conception that inquiry and modeling develops the academic skills of all students through active learning and the construction of new knowledge using students' prior experiences and misconception. Providing evidence that inquiry and modeling "work" for all students supports teachers in the adoption of Physics First by addressing the "practicality ethic of teachers" (Altrichter, 2005). Qualitative studies have demonstrated that teachers who do not believe their students are capable of or prepared for learning the science they are teaching make instructional decisions that lower the quality of instruction (Gilbert & Yerrick, 2001; Prime & Miranda, 2006). Conversely, teachers who hold high expectations for their students increase the participation of historically underserved students in science (Luft, da Cunha, & Allison, 1998).

Now that I have reviewed the literature on the OTL gap that disproportionately impacts historically underserved students and how a Physics First approach addresses these gaps, I examine the methodological literature on the effects of a Physics First approach on improving student learning in science.

Review of the Methodological Literature

A relatively small number of quantitative studies of Physics First have been completed and only those focused on student achievement in science or math are included here. Studies concerned that assessed changes in teacher pedagogy or gathered data through student surveys are excluded. The remaining studies are all causal-comparative, quasi-experimental designs. Two comparison designs are common. Students either self-select into either a traditional or Physics First science sequence and

results of the two groups are compared or all students are placed in the same course sequence during a period of time and results are compared between one or more cohorts experiencing each sequence. A variety of measures of student achievement (e.g., teacher constructed pre- and post-tests, state tests, college-readiness exams, AP course enrollments and test scores) have been employed in these studies. The use of inferential statistics to detect group differences (e.g., ANOVA, t-tests, and chi-square tests) is common. Disaggregation of results by gender is not uncommon, but reporting of results for students from historically underserved groups in these studies is non-existent. I now summarize the results of the ten studies of the effects on student achievement of inverting the high school science sequence with or without modeling instruction.

O'Brien and Thompson (2009) investigated physics performance of ninth graders and twelfth graders in seven high schools in Maine. The 321 students formed five distinctive groups of students in this study: (a) ninth-grade students who experienced traditional instruction, (b) ninth-grade students who experienced modeling-based instruction, (c) ninth-grade honors-level students who experienced traditional instruction, (d) ninth-grade honors-level students who experienced modeling-based instruction, and (e) twelfth-grade students who experienced traditional instruction. A 27-item multiple choice survey developed using items from three established instruments (including the Force Concept Inventory), served as the pre- and post-tests. Among all five groups, pre-test results showed little understanding of concepts in kinematics and mechanics. The post-test scores indicated the honors-level ninth graders had the highest normalized gain between the pre- and post-tests scores regardless of whether or not modeling instruction

was used. However, for non-honors 9th graders, the students experiencing modeling instruction had six times the normalized gain scores of that of the non-honors students who did not have modeling instruction.

Liang et al. (2012) conducted a causal-comparative study of Physics First with modeling in two predominantly White, middle class high schools in the Mid-Atlantic. The Force Concept Inventory (FCI) was used as a pre- and post-test in two comparisons. Liang, et al. (2012) used analysis of covariance to compare the mean pre- and post-test scores for groups of students on the FCI, controlling for differences in groups on the pre-test. For the comparison of mean scores on the FCI, 9th grade honors physics course with modeling and the 12th grade honors physics course without modeling, students in the honors physics course with modeling had significantly greater mean scores on the FCI ($p < .001$, effect size = 2.45) after controlling for pre-test scores. A comparison of students in non-Physics First courses found similar results on the FCI for students in courses employing modeling compared to non-modeling courses ($p < .001$, effect size = 2.62).

Bermudez (2014) examined the effects of transitioning from a biology-chemistry-physics course sequence to a biology-physics-chemistry sequence (“physics second”) at a high-poverty, predominantly Latino public high school in California. Eight years of state end of course exam results in physics and chemistry (four years for each sequence) were analyzed using independent t-tests for the means and chi-square tests for the proficiency level of the student. A 2 x 2 factorial ANOVA was employed to detect differences in mean scores based on gender. Bermudez found that student achievement, as measured by mean scores and proficiency level, on the end-of-course chemistry exam was

significantly greater ($p < .001$) for students in the “physics second” sequence, but was significantly lower ($p < .001$) on the end-of-course physics exam. Gender differences in mean scores unrelated to course sequence were observed on the physics end-of course exam, although no gender differences were found for chemistry. A significant limitation of this study noted by the researcher was the change in graduation requirements from two to three science courses that occurred concurrently with the implementation of the “physics second” sequence. As a result of this change, the number of students taking the physics end-of-course exams increased eight-fold between the two cohorts. A measure to determine the academic equivalence (or lack thereof) of the two groups would benefit efforts to interpret Bermudez’s findings (Isaac & Michael, 1995; Johnson, 2001).

Mary (2015) examined the effect of science course sequence on student performance on annual end-of-course state science and math assessments at two large, diverse, suburban public high school in Texas. Three cohorts of students (9th graders in 2011, 2012, and 2013) who self-selected into a traditional sequence or a Physics First sequence were used. However, state testing requirements in 2014 eliminated end-of-course assessments in chemistry, physics, and geometry so only the students in the traditional sequence in the first cohort provide data across all three years of either sequence, limiting Mary’s ability to investigate the impact of the full science course sequence. Mary reports that teachers were required to use the same instructional materials, district curriculum guides, resources, and common assessments regardless of which grade the course was taught. To control for differences between groups due to student self-selection, Mary employed exact matching for gender, at-risk status, high

school, and ninth grade math course. Race/ethnicity was found not to be a covariate for matching. Regression was used to combine 8th grade math and 8th grade science performance into a single principal component score and included in the matching process. Employing t-tests of end of course means and ANOVA, Mary found there was not a statistically significant effect of science course sequence on student performance on end-of-course assessments in science or math.

To explore the effects of a Physics First sequence on math achievement, Glasser (2012) used the scores from state end-of-course eighth grade math assessment as baseline data on six classes of students; three of which began the traditional course sequence (graduating classes of 2000-2002) and three in a Physics First sequence (graduating classes of 2003-2005). Students in all six cohorts in a private school in Pennsylvania were found to be equivalent in quantitative reasoning skills at the end of eighth grade using a chi-square test. In the fall of tenth grade all students took the PSAT exam. Using a t-test, Glasser found a statistically significant difference ($p < .01$) between the means of math reasoning percentiles favoring the last two of the three graduating classes enrolled in the Physics First sequence compared to the pooled mean percentile of the three classes enrolled in a tradition sequence of biology, chemistry, and physics.

Bouma (2013) also reports increased math achievement for students in a Physics First sequence compared to those in a traditional science sequence in an urban, majority-minority, private, all-boys, college-preparatory high school on the West Coast. Ex post facto measures from standardized math tests (High School Placement Test [HSPT] and SAT) at the school site were used to determine the math achievement of two groups of

students in the same graduating class (cohort) over their high school career. The ninth grade science course determined the two groups: those that took ninth-grade physics (PF) and those that took no science in the ninth grade (non-PF). Controlling for prior math achievement, race/ethnicity, and SES using ANCOVA, Bouma found students in the second PF cohort scored significantly higher on SAT math scores than non-PF students ($p < .05$ and effect size = .27), but there was no significant difference between the two groups in the first cohort.

Williams (2009) examined the math and science achievement of three cohorts of students (graduating classes of 2007, 2008, and 2009) in a racially-diverse suburban high school in Illinois. In this study, students self-selected into a Physics First science sequence or a traditional science sequence beginning with biology. Students were then placed into an honors or regular section of each course based on their 8th grade EXPLORE scores and math placement. Academic achievement was measured using the 9th grade EXPLORE, 10th grade PLAN, and 11th grade ACT for both science and math and the state's science test at grade 11. EXPLORE, PLAN, and ACT measure students' scientific reasoning skills and mathematical computation and reasoning skills while the state science test measures content knowledge and skill application. A one-way, between-groups analysis of variance (ANOVA) compared the mean score for each achievement test. An independent samples t-test identified between-groups differences on the state science test. Williams (2009) determined that gains in science student achievement from the 9th grade EXPLORE to the 11th grade ACT from grade 9 to grade 11 varied significantly by course ($p < .001$). Gains were greater for honors and regular Physics First

students than for the honors and regular biology students. Honors biology students and honors Physics First students demonstrated similar levels of content knowledge and skills on the state science test. There was also no significant difference in the mean scores of regular biology students and regular Physics First students on the state science test.

Williams also found that science course sequence did not have a significant effect on mathematics achievement or growth, but there was a significant difference ($p < .001$) between honors students and regular students in terms of the amount of growth between the 9th grade EXPLORE and 11th grade ACT. Using two-way ANOVA, Williams found that female and male students exhibited similar achievement scores and growth over time in both math and science regardless of the science course sequence.

Dye et al. (2013) also examined the effects of transitioning from a biology-chemistry-physics course sequence to a physics-chemistry-biology sequence and then incorporation of modeling instruction in the inverted science sequence over an eight year period at a southeastern, urban, Catholic high school. The first four cohorts were traditional instruction in the biology-chemistry-physics sequence (control group) followed by two cohorts of students taking a Physics First science sequence with traditional instruction (treatment 1). The last two cohorts are students taking a Physics First sequence with modeling instruction (treatment 2). Gain scores were calculated for each group by subtracting the PLAN science mean score (administered in fall of the 9th grade year) from the 11th grade ACT science mean scores. Gain scores were largest for the students in the Physics First sequence with modeling (4.3) compared with a gain of 4.1 for students in Physics First with traditional instruction and a gain of 2.8 points for

the control group. Cohen's d for the mean ACT scores reported by Dye et al. was .29 for the students in the Physics First sequence with modeling compared to the control group and .15 compared to Physics First with traditional instruction (treatment 1). Changes in the percentage of students who were college- and career-ready in science were also reported with the gains for students in the Physics First sequence (20.8% for students experiencing modeling instruction and 17.8% for students experiencing traditional instruction) compared to a gain of 2.5% for students in the control group. Inferential statistics are not provided for the changes in percentage of students who are college- and career-ready in science.

Goodman and Etkina (2008) investigated the benefits of teaching a mathematically rigorous ninth-grade physics course based on algebra alone. Topics for the ninth-grade physics course were drawn from the AP Physics B curriculum and the new science sequence was implemented in a New Jersey county vocational/technical high school founded in 1999. The study focused on the number of students taking AP exams at the school compared to the average for the state as well as the number of students receiving scores of 3 or higher (considered "passing") compared to the state. After four years of implementing this mathematically rigorous Physics First approach, students at the school took the AP Physics B exams at a rate 14 times that of the state and the percentage of students passing the exams was also 14 times higher than the state. Goodman and Etkina also report that the mean number of science courses completed by students in the school rose from 3.4 to 4.2 over the same four year period.

Gaubatz (2013) examined the effects of transitioning from a biology-chemistry-physics course sequence to a “modified” physics-chemistry-biology sequence at a diverse, suburban public high school in the Midwest. The modified physics course for freshman was either an Honors Physics course for students enrolled in geometry or other higher-level math or a “GeoPhysics” course. Results from eight cohorts of students, four from each course sequence were compared in this program evaluation. Enrollment in honors or AP courses increased significantly for students in the Physics First sequence (using a two-tailed t-test, $p < .05$ for freshmen, and $p < .01$ for sophomores and juniors), AP Biology enrollment tripled, and enrollment in AP Environmental Science, AP Chemistry, and AP Physics C doubled. Student growth in science as measured by increases in mean scores from the 8th grade EXPLORE to the 11th grade ACT was not statistically significant between the “modified” physics-chemistry-biology and the traditional sequence cohorts.

Overall, the results of these ten studies on the effects of inverting the high school science sequence on student achievement in math and science are ambiguous. Incorporating modeling instruction into physics significantly improves student achievement on the Force Concept Inventory (Liang et al., 2012; O’Brien & Thompson, 2009,). Both Gaubatz (2013) and Goodman and Etkina (2008) report a Physics First approach increases student enrollment and achievement in advanced science coursework. However, on more standardized measures of science (end-of-course exams), inverting the science sequence does not appear to improve student achievement (Bermudez, 2014; Mary, 2015). The evidence of the effects of implementing a Physics First approach on

math and science achievement and growth is also mixed with some researchers reporting improvement in achievement and others reporting no effect (Bouma, 2013; Dye et al., 2013; Gaubatz, 2013; Glasser, 2012; Williams, 2009). None of these studies report results for historically underserved students.

To determine if the science college readiness of historically underserved students can be improved by implementing a Physics First science sequence with modeling, a retrospective cohort observational study using multiple linear regression and logistic regression was conducted. Change in science college- and career- readiness and interest in STEM careers between grades 8 and 11 will be examined for four successive cohorts of students. The first two cohorts will be students experiencing a traditional science sequence and the two subsequent cohorts will be students experiencing a Physics First with modeling approach. Unlike the studies examined here, results will be disaggregated for students from historically underserved groups.

There is major disagreement among authors of educational research texts as to the status of correlational design compared to causal-comparative designs (Johnson, 2001). Although many authors treat causal-comparative designs as superior to correlational designs, Johnson argues that both approaches are on an equal footing. In both causal-comparative and correlational designs, some evidence of causality can be obtained by identifying potential confounding variables and attempting to control for them (Johnson, 2001). I used 8th grade science college- and career- readiness test scores and interest in STEM to control for differences in prior achievement and interest between the cohorts

and thereby reduce the plausibility of rival explanations (Campbell & Stanley, 2015; Johnston, 2001; Krathwohl, 2009)

Summary of the Research Literature and Application to the Study

A Physics First science sequence for all students increases the access of historically underserved students to challenging science curriculum and high-quality instruction. Incorporating inquiry and modeling develops the academic skills of all students through active learning and the construction of new knowledge using students' prior experiences and misconception. While some aspects of the Physics First approach described here are culturally responsive, the cross-cultural and socio-political perspectives of critical race theory are largely ignored. Further, many will argue that schools alone cannot solve the achievement gap due to the pernicious effects of societal factors affecting the lives of students of color and families who are economically disadvantaged. Nonetheless, I join many other voices for equity in calling on our schools to do more to close the opportunity gaps experienced by our historically underserved students. Overall, the review of the literature presented here suggests that using a Physics First approach for high school science has the potential to close opportunity to learn gaps experienced by historically underserved students in all three dimensions of OTL: content coverage, content exposure and emphasis, and instructional delivery. Closing these opportunity to learn gaps is a moral obligation holding the promise of increasing the science college readiness of these students and their preparation for STEM careers. Given the apparent absence of evidence of the effects of a Physics First approach on improving student achievement of historically underserved students in science, this

proposed quantitative study will contribute to the knowledge base of effective high school science practices (Kratwohl, 2009). Specifically, the study will determine if, compared to their peers in a traditional science sequence, historically underserved students completing a Physics First science sequence have higher science achievement gains over time, higher rates of science college- and career-readiness, and greater interest in STEM. I outline the study setting and participants as well as the instruments and methods for answering these questions next.

CHAPTER 3 METHODOLOGY

While there are many factors contributing to disparate educational outcomes for low-income and Black and Hispanic/Latino students, the purpose of this study was to determine if the science college readiness of historically underserved students can be improved by implementing a district-wide, inquiry-based high school science sequence comprised of coursework in physics, chemistry, and biology. This Physics First approach to high school science has the potential to close opportunity to learn gaps experienced by historically underserved students in all three dimensions of OTL: content coverage, content exposure and emphasis, and instructional delivery. Gaps in content coverage experienced by historically underserved students are closed by ensuring all students are exposed to the same content in physics, chemistry, and biology. Incorporating inquiry and modeling develops the academic skills of all students through active learning and the construction of new knowledge, closing gaps in content exposure and emphasis. Ensuring all science teachers have strong content and pedagogical knowledge, belief in their own abilities to teach all students, and belief that their students are capable of learning science concepts leading to science college-readiness closes gaps in instructional delivery. A retrospective cohort observational study (Hoffmann & Lim, 2007; Mann, 2003) was used to address the primary research question: do historically underserved students in a Physics First science sequence have 1) higher science test scores, 2) higher rates of science college- and career-readiness, and 3) greater interest in STEM careers in grade 11 compared to their peers in a traditional science sequence?

Research Methods

Because the purpose of the study was to examine the relationship between the implementation of Physics First and the science college readiness of historically underserved students, a quantitative research design was appropriate (Krathwohl, 2009). A quantitative approach was also warranted since the three facets of the research question are predetermined, narrow and specific, and can be answered using quantifiable data in an objective manner (Plano Clark & Creswell, 2015). In contrast, qualitative studies seek to understand what is going on in a particular setting or with participants (Maxwell, 2013). Given the relatively small number of studies that have examined the effects of implementing Physics First on student achievement, the inconsistent findings from these studies, and the lack of results reported for students from historically underserved populations, I argue that the first order of business is to determine the effects of implementing Physics First on the science college readiness of historically underserved students. Without first ascertaining whether such an approach yields benefits, understanding *how* and *why* such an approach does or does not close science college- and career-readiness gaps is putting the cart before the horse. Qualitative studies to better understand how and why a Physics First approach does or does not work would be valuable follow-up studies (Maxwell, 2004).

This study was a retrospective cohort observational design using pre-formed groups (Hoffmann & Lim, 2007; Mann, 2003); random assignment of students to the Physics First science sequence was neither feasible nor educationally justifiable (Cochran, 1983; Osbourne, 2008; Plano Clark & Creswell, 2015). Offering a traditional

science sequence alongside a Physics First science sequence in each school and randomly assigning students to each sequence would have required adoption of instructional materials aligned with the NGSS for each sequence, as well as two strands of teacher professional development. In addition, garnering educator and parental support for random assignment is often difficult as it is a forced choice. Campbell and Stanley (2015) promote the consideration of single-group experiments when random assignment is not possible. A matched-pair randomized control trial of schools within the district would present similar challenges as a randomized control trial at the classroom level, in addition to the limited number of schools (10) available for forming matched-pairs (Ji, DuBois, Flay, & Brechling, 2008). Campbell and Stanley (2015) and Krathwohl (2009) assert that quasi-experimental designs done well can provide evidence for policy decisions when a true experimental design is not feasible. They would also agree that, compared to a randomized experiment, a major weakness of quasi-experimental designs is the difficulty in eliminating rival explanations.

I used a retrospective cohort observational study (Hoffmann & Lim, 2007; Mann, 2003) with two groups consisting of two 9th grade cohorts each. In this design, the cohorts were naturally formed based on year of enrollment in 8th grade. Pre- and post-measures were administered to each cohort. Campbell and Stanley (2015) assert a number of threats to internal validity are more effectively controlled when assignment to groups are similar (e.g., students do not self-select, assignment is not based on previous performance) and pre-measures are used to confirm similarities of the groups or control for initial differences between the groups. Nonetheless, a major confounding variable

(Krathwohl, 2009) that could not be controlled is the variability in science instruction students receive during the 8th grade after the November administration of the pre-test. Other possible confounding variables are due to the use of four successive annual cohorts and include differences in instructional time (i.e., length of calendar, loss of instruction due to inclement weather), differences in class size, and changes in science teachers each year. I argue that these confounds are unlikely rival explanations. Further, the inclusion of all students from naturally formed groups enrolled for three years in high school eliminates selection bias. I now discuss both the participants and the pre- and post-measures used in the study.

Participants

The study was situated in a suburban school district of 30,000 to 50,000 students in the northwest United States with approximately equal numbers of students of color and White students. Enrollment in each of the district's 10 high schools ranged from 200 to 2200 students, with more than one-third of students qualifying for free or reduced price lunch. The district in this study implemented a Physics First model for all students in 2012; thus, a randomized control trial was not possible (Cochran, 1983; Issac & Michael, 1997; Krathwohl, 2009). Student experiencing the traditional science sequence entering 9th grade in the fall of 2010 and in the fall of 2011 served as the control group in this retrospective observational study. Freshmen experiencing an inquiry-based science sequence consisting of physics, chemistry, and biology with modeling entering 9th grade in the fall of 2012 and in the fall of 2013 were the treatment group. The use of two cohorts in the control and treatment groups increased the trustworthiness of the results

(Krathwohl, 2009) and mitigated any possible implementation dip during the first year of curriculum change (Fullan, 2002).

In order to identify differences in outcomes between the two science sequences, only students enrolled on May 1 for three consecutive years in high school were included in the analysis. Science test scores and STEM interest in grade 8 were used to control for any initial differences between the two groups (Tuckman, 1994). Krathwohl (2009) notes that dramatic effects are uncommon in education; therefore, researchers should “design studies with sensitivity sufficient to establish weak effects” (p. 228). Including all students, as opposed to randomized or stratified sampling, increased the certainty of inference and the power of inferential statistics produced for specific student groups (e.g., Hispanic/Latino students) within the historically underserved student category (Briggs, 2008; Field, 2016; Krathwohl, 2009). Larger sample sizes also increased the reliability of effect sizes (Slavin, 2008; Slavin & Smith, 2009)

Procedures

For each of the four cohorts, 9th grade students entering in the fall of 2010 and 2011 experiencing a traditional science sequence and 9th grade students experiencing a Physics First science sequence entering in the fall of 2012 and 2013, the EXPLORE science test was administered to 8th grade students during a three week window in November. The 9th grade EXPLORE, 10th grade PLAN, and 11th grade ACT were administered on a single day in the spring. The 9th grade EXPLORE was administered in April for all four cohorts. The 10th grade PLAN was administered in April for the 2010, 2011, and 2012 cohorts and in March for the 2013 cohort. The 11th grade ACT was

administered in April for the 2010, 2011, and 2013 cohorts and in March for the 2012 cohort. Figure 3 summarizes the cohorts and testing dates. All other conditions for the administration of the EXPLORE, PLAN and ACT were identical across all four cohorts, observing time and testing conditions specified for each test.

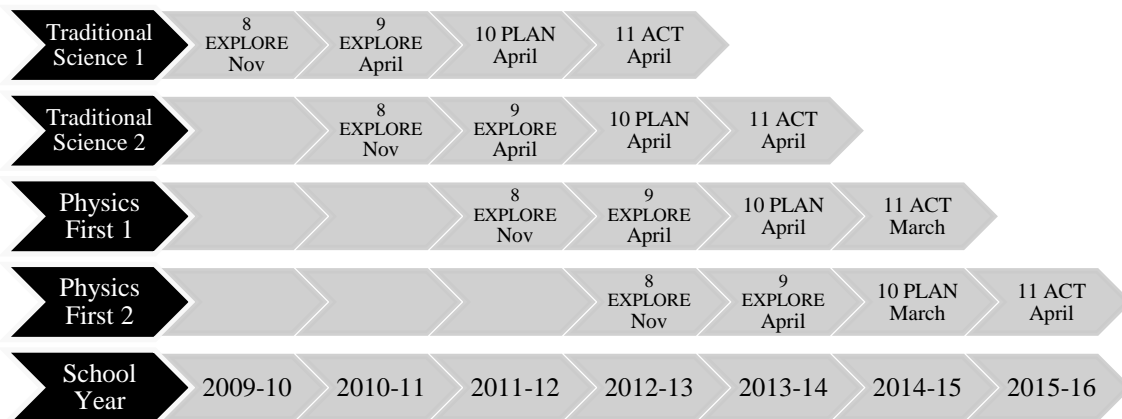


Figure 3. Science college readiness testing by cohort.

Because this retrospective observational study relied on extant data from the district, I first obtained approval from the district's Research Committee to receive electronic files of de-identified data to conduct the study. The district's criteria for approving applications to conduct research include IRB review and approval. The requested files, one for each the four cohorts of students, consisted of all 8th grade students enrolled on the first school day in May joined with the electronic files provided by ACT with results of college-readiness testing for each year in grades 8 – 11. In this study, the group membership of a student is based on the race/ethnicity, eligibility for free or reduced-price school meals, and English Language Learner status on May 1 of the student's 9th grade year. I requested that the district remove from the files for each cohort

any and all student identifiable information (i.e., student identification numbers, names, birthdates, addresses, telephone numbers) originating from any of the five files merged into the file for the cohort. These four merged and de-identified cohort files of extant data were used to answer the research questions.

The four electronic cohort files were then imported into SPSS statistical software. In the SPSS files, I calculated fields to aid in answering the research questions. I assigned a subject number to each student in the four files and a variable designating which of the four cohorts the student was a member. Dummy variables were created for each race/ethnicity from the single race/ethnicity field in the files supplied by the district. An additional variable designating if the student was a member of a historically underserved population was calculated. Historically underserved students are Native American/Alaskan Native, Black, Hispanic/Latino, Pacific Islander/Native Hawaiian, English language learners, and economically disadvantaged students. Variables to designate whether a student met the science college- and career-readiness benchmark were calculated for each assessment. A field was also calculated to indicate if a student had an expressed or measured interest in STEM on the ACT Interest Inventory at the time of each assessment.

All source files and the SPSS files are password protected and stored on a password-protected computer on a password-protected network. Copies of the data sets are securely stored on an electronic storage device in a locked filing cabinet and securely stored remotely on a password protected server. The files will be retained for three years after completion of the study and then destroyed.

Instruments and Measures

The EXPLORE, PLAN, and ACT assessments are comprised of a standardized, curriculum-based battery of multiple-choice tests in reading, math, English, and, most importantly for this study, science (ACT, 2014). Results from these four tests along with the non-cognitive component of each of the assessments, the ACT Interest Inventory, “help students plan for further education and explore careers, based on their own skills, interests, and aspirations” (ACT, 2014, p. 2). The ACT is tied more closely to high school curricula than the SAT, is grounded in reviews of state content standards and periodic national curriculum surveys, and emphasizes content mastery over test-taking skills (Atkinson & Geyser, 2009). Nonetheless, Atkinson and Geyser (2009) note that the ACT lacks the depth of subject matter coverage that one finds in other achievement tests such as AP exams or SAT subject tests, citing the science test as an example. Instead, the ACT science test emphasizes understanding the practices and process of science rather than specific science content (Schultheis & Kjervik, 2015; Williams, 2009). Content specifications for the ACT science test are included in Appendix A. The ACT EXPLORE, PLAN, and ACT assessments provide data on student growth over time in science reasoning, science college-readiness, and interest in STEM careers, all of which are central to the research question of this study. The use of PLAN and ACT scores for program evaluation is supported by both the psychometric properties and content validity of the tests (ACT, 2014). For fifteen years, the *Standards for Educational and Psychological Testing* authored by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in

Education (1999) opposed the use of college admissions tests for program evaluation.

“Admission tests, whether they are intended to measure achievement or ability, are not directly linked to a particular instructional curriculum and, therefore, are not appropriate for detecting changes in middle school or high school performance” (American Educational Research Association et al., 1999, p. 143). However, Slavin (2008) argues that nationally standardized tests can be used to assess differences in performance between two groups and may be more accurate measures because they are not directly linked to a curriculum or instruction received by one group and not the other. Slavin’s argument appears to have prevailed. The most current rendition of the *Standards for Educational and Psychological Testing* now states that a variety of tests can be used for evaluating programs, including standardized achievement tests (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The language in the previous version of the *Standards* asserting that the use of admission tests for program evaluation was inappropriate appears to have faded away.

When tests are used to evaluate a program or policy, evidence of the validity of the use of test scores for that purpose should be provided (American Educational Research Association et al., 2014). “Validity refers to the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests” (American Educational Research Association, 1999, p. 9). Arguments for the validity of an intended inference made from a test usually combine logical, empirical, and/or theoretical sources (ACT, 2014; Krathwohl, 2009). Validity evidence for the ACT, the

College Readiness Benchmarks, and the Interest Inventory are described in the remainder of this section along with the instruments themselves.

ACT college and career readiness science tests. ACT's college and career readiness tests are curriculum-based and measure what "students are able to do with what they have learned in school, not abstract qualities such as intelligence or aptitude" (ACT, 2014, p.1). The three tests are scored along a common scale extending from 1 to 36; the maximum score on ACT Explore (for students in grades 8 and 9) is 25, the maximum ACT Plan (grade 10) score is 32, and the maximum ACT score is 36 (ACT, 2014). The standard error of measurement is approximately 2 scale score points for each of the subject-area test scores (ACT, 2014). ACT equates test forms across years so that "scale scores are comparable across test forms and test dates" (ACT, 2014, p.51).

The ACT Science Test is a 40-item, 35-minute test that measures the interpretation, analysis, evaluation, reasoning, and problem-solving skills required in the natural sciences (ACT, 2014). The content of the Science Test is drawn from biology, chemistry, physics, and Earth/ space science and assumes students are both familiar with the nature of scientific inquiry and have been exposed to laboratory investigation (ACT, 2014). "As with the NGSS, the ACT science readiness scores clearly emphasize the importance of understanding the practices and process of science rather than the memorization of facts" (Schultheis & Kjervik, 2015, p.25).

The PLAN Science test is a 30-item, 25-minute selected-response assessment that calls on students to critically examine information and possible interpretations and draw conclusions or make predictions. Content of the test is typically covered in early high

school general science courses and is drawn from the biological sciences, earth/space science, physics, and chemistry. Scientific reasoning skills are emphasized over recall of specific scientific content (ACT, 2013b).

The EXPLORE Science test is a 30-item, 25-minute selected-response assessment measuring scientific reasoning skills acquired up to grade 8. The content of the test is typically covered in science courses through grade 8 related to life science, Earth/space science, and physical science. Like the PLAN assessment, EXPLORE emphasizes scientific reasoning skills (e.g., drawing conclusions, making predictions) over recall of specific scientific content (ACT, 2013a).

Validity Evidence for ACT Test Scores. Examining the first year college success rates of over 190,000 students at 192 institutions, Sawyer (2010) found that both high school GPA and the ACT composite score predict academic success in the first year of college, with the ACT composite score a better predictor of higher GPAs (i.e., 3.5 and above) in the first year of college. After examining the performance of nearly 190,000 first-time freshmen at four-year colleges and universities, Westrick, Le, Robbins, Radunzel, and Schmidt (2015) conclude that ACT Composite scores are highly correlated with first year academic performance across a range of institutions from selective to open enrollment. In contrast, Bettinger, Evans, and Pope (2011) found that after controlling for selectivity of enrolled college, high school GPA, race/ethnicity and gender, and college major, the ACT science test score was not correlated with first year overall college GPA, second year overall college GPA, or persistence in college (although the math and English test scores were predictive). Based on their findings, Bettinger et al.

argue that since the composite score includes the non-predictors of the ACT reading and science scores, using only the ACT math and English test scores may better identify students who will be successful in college.

ACT College Readiness Benchmarks. The ACT Science College Readiness Benchmark of 23 is the minimum ACT test score required for students to have a high probability of success in a college biology class (ACT, 2013d). Corresponding Science Benchmarks for EXPLORE (18 at grade 8; 19 for grade 9) and PLAN (20 in grade 10) gauge student progress in becoming ready for studying science in college (ACT, 2013d). Students who meet the science benchmark on the ACT have approximately a 50% chance of earning a B or better and approximately a 75% chance or better of earning a C or better in college biology (ACT, 2014). Students who meet the benchmark on the EXPLORE or ACT PLAN science tests have approximately a 50% chance of meeting the ACT Benchmark in science, and are likely to have approximately this same chance of earning a B or better grade in college biology by the time they graduate high school (ACT, 2014). After examining freshman college math grades and ACT math benchmark scores of Minnesota students, Maruyama (2012) concluded that the math ACT college readiness benchmark is a more accurate threshold for earning a B or better grade than the C or better grade. Maruyama also suggests college readiness results from the ACT would be more usable if probabilities of success were provided to students at every score point, not just for the college readiness benchmark. For example, a student with an ACT science score of x has a 40% chance of attaining a grade of B or higher in college biology.

ACT College Readiness Benchmarks are empirically derived based on the actual performance of students in college. Data from 214 institutions and over 230,000 students were used to establish the benchmarks (ACT, 2014). The sample of colleges is weighted by ACT so that it is representative of all ACT-tested college students in terms of college type (2-year and 4-year) and selectivity (ACT, 2014). The College Readiness Benchmarks for EXPLORE and PLAN were developed using records of students who had taken EXPLORE or PLAN, followed by the ACT in grades 11 or 12 (ACT, 2014). Separate benchmarks were developed for EXPLORE for grade 8 and for grade 9 (ACT, 2014). The sample sizes used to develop the EXPLORE and PLAN Benchmarks ranged from 210,000 to approximately 1.5 million students depending on the test (ACT, 2014). To establish the benchmarks, the probability of meeting the appropriate ACT Benchmark was estimated for each EXPLORE and PLAN score (ACT, 2014). The EXPLORE and PLAN science test scores corresponded most closely to a 50% probability of meeting the benchmark for science on the ACT (ACT, 2014).

Validity Evidence for ACT College Readiness Benchmarks. Using logistic regression, Allen and Sconing (2005) established readiness benchmarks for common first-year college courses based on ACT scores. These benchmarks for the EXPLORE, PLAN, and ACT were updated in 2013 using data from more recent high school graduates (Allen, 2013). Based on Allen's analysis of over 40,000 students from 90 colleges, the science college readiness benchmark was decreased by one point on each assessment. These updated 2013 benchmarks were applied to all science test scores in this study. Noble, Davenport, Schiel, and Pommerich (1999) used stepwise multiple

regression to investigate the extent to which non-cognitive characteristics explained differential ACT performance of racial/ethnic and gender groups. The researchers found that 36% of the variability in ACT science scores was attributable to specific coursework taken and grades earned in high school, while less than 3% of variance in ACT science scores was related to gender or race/ethnicity (Noble et al., 1999). McNeish, Radunzel, and Sanchez (2015) replicated these findings. However, the findings related to race and ethnicity are only for Black students due to the small number of Latino and Native American students in the study's data set. In a study of 190,000 ACT-tested students enrolling in college as first-time students in fall of 2000 through 2006, Radunzel and Noble (2013) found that ACT Benchmark scores overestimated the chances of success for students of color in college degree attainment, but less so than the use of high school GPA. The authors also found that ACT benchmark scores "slightly over-predicted students' chances of progressing towards and completing a degree for lower-income students" (Radunzel & Noble, 2013, p. 41).

ACT Interest Inventory. The ACT Interest Inventory (UNIACT) helps students explore personally relevant career options (both educational and occupational) during high school (ACT, 2014). UNIACT results are reported for six scales paralleling the six interest and occupational types in Holland's theory of careers (ACT, 2014). Scale names (and corresponding Holland types) are Science & Technology (Investigative), Technical (Realistic), Administration & Sales (Enterprising), Arts (Artistic), Business Operations (Conventional), and Social Service (Social) (ACT, 2014). Each scale consists of work-relevant activities (e.g., build a picture frame, conduct a meeting, help settle an argument

between friends) familiar to students, either through participation or observation (ACT, 2014). Two work task dimensions underlie Holland's six interest and occupation types: 1) working with data versus ideas and 2) working with things versus people (Prediger & Swaney, 2004). The term *data* refers to working with numbers, files, accounts, or business procedures while working with *ideas* is about forming insights, theories, new ways of saying or doing something (ACT, 2009a). The term *people* refers to helping, serving, informing, caring for, or selling things to people as opposed to working with *things* (e.g., machines, tools, living things, and materials) (ACT, 2009a). Students respond to 72 items on the UNIACT using a three-choice response format (*dislike, indifferent, like*) (ACT, 2014). Nationally representative norms for grades 8, 10, and 12 are based on a nationally representative sample of 257,567 students from 8,555 schools. (ACT, 2009a; ACT, 2014). Internal consistency reliability coefficients for the six 12-item scales range from .84 to .91 (ACT, 2014).

Validity Evidence for the ACT Interest Inventory. A number of studies have confirmed the criterion-related validity and structural validity of the ACT Interest Inventory (ACT, 2009a). Evidence of criterion-related validity occurs when individuals with the same occupational choice, college major, or occupation express interest that would assign them to the same career cluster in the inventory (ACT, 2009a). In a study of nearly 11,000 high school seniors who indicated they were very sure of their career choice, 42% were assigned to the same career cluster based on their expressed interests (ACT, 2009a). Principal component analysis has been used to confirm that the data versus ideas and things versus people work task dimensions underlie the six ACT Interest

Inventory scores (ACT, 2009a). After analyzing three databases providing a wide range of perspectives on basic worktasks, Prediger and Swaney (2004) confirmed that the data versus ideas and things versus people work task dimensions underlie diverse types of occupational data. Day, Rounds, and Swaney (1998) examined factor loadings on the data versus ideas and people versus things work task dimensions for racial/ethnic groups and concluded that “the ACT Interest Inventory has validity for use with diverse racial/ethnic groups in the United States” (ACT, 2009a, p. 7).

STEM Interest. A student is classified as having interest in STEM if, on the ACT Interest Inventory, the student plans a STEM major or occupation following high school (*expressed STEM interest*) or the student had a highest ACT Interest Inventory score in Science & Technology or had a highest ACT Interest Inventory score in Technical and a second-highest score in Science & Technology (*measured STEM interest*) (ACT, 2015). *Science & Technology* are work tasks that involve “investigating and attempting to understand phenomena in the natural sciences through reading, research, and discussion” (ACT, 2009a, p. 4). *Technical* work tasks involve “working with tools, instruments, and mechanical or electrical equipment. Activities include designing, building, repairing machinery, and raising crops/animals” (ACT, 2009a, p. 4). On the ACT, a student identifies a major or occupation from a comprehensive list. For the interest inventories accompanying the EXPLORE and PLAN assessments, students are asked to identify one of 26 career areas the student is most interested in (ACT, 2009a). Five of these career areas – Computer/Info Specialties, Engineering & Technologies, Natural Science & Technologies, Medical Technologies, and Medical

Diagnosis & Treatment – are “the closest you can get to ACT’s STEM occupation titles”, encompassing 83 of the 93 STEM careers (Kyle Swaney, personal communication, October 6, 2016).

Role of the Researcher

My experience in education includes teaching mathematics at the high school level as well as administrative experience at a state education agency and in a local education agency. My administrative experience includes supervising the administration of standardized testing, reporting of testing results for accountability and improvement, supplying data to inform the selection of instructional materials, and conducting program evaluations. I have not been involved in the selection of science curriculum or teacher professional development related to science curriculum implementation nor do I administer college readiness assessments nor instruct high school students in science. For the past fifteen years, a substantial portion of my duties has been the compilation, examination, and analysis of data which repeatedly reveal the disproportionate outcomes in student achievement and graduation rates for students of color and students from low-income families. I intentionally began my dissertation with the powerful assertion by Henry Levin (2009) that “educational equity is a moral imperative for a society in which education is a crucial determinant of life’s chances” (p. 5). My commitment to educational equity, the moral imperative I share with Mr. Levin, is a source of potential bias toward positive outcomes for historically underserved students in this study. My use of extant data in a retrospective observational study greatly reduced researcher expectancy bias (Krathwohl, 2009; Mann, 2003). Further, I took several steps to increase

the integrity of the results and diminish researcher bias including: 1) specifying the research question in advance of the study, 2) identifying in advance data to answer the research question, 3) applying statistical techniques to analyze the data, and 4) specifying significance levels for inferential statistical tests in advance of analyzing the data. I now describe the statistical analysis used to answer each part of the research question in this pre-planned study.

Data Collection and Analysis

To reduce plausible rival explanations, statistical methods were employed to adjust for initial differences in science achievement between the two groups (if present) and reduce the effect of unwanted variables (Krathwohl, 2009). Science test scores, science college readiness, and STEM interest in the 8th grade year were used to control for any initial group differences (Tuckman, 1994). Statistical controls for initial differences work well when initial differences between treatment and control groups are small (Slavin, 2008). Inferential statistics were computed using the conventional significance level (α) of 0.05 for social science research to detect if there are differences between the populations (Bloom, Hill, Black, & Lipsey, 2008). A significance test does not tell the *size* of a statistical difference between two measures, but effect size does (Bloom et al., 2008; Fields, 2016). Effect size can be interpreted in various ways and the method selected should be determined by the research question (Fields, 2016). The effect size represents the magnitude of an intervention in statistical terms, specifically in terms of the number of standard deviation units by which the treatment group outperforms (or underperforms) the comparison group (Fields, 2016). The commonly used interpretation

suggested by Cohen classifies effect sizes as small ($d=0.2$), moderate ($d=0.5$), and large ($d=0.8$) (Bloom et al., 2008; Fields, 2016). Bloom et al. assert that there is no reason to believe Cohen's rule of thumb applies to the effects of educational interventions or, more specifically, to effects on the standardized achievement tests widely used as outcome measures" for such interventions (p. 295). Bloom et al. calculated average gains in effect size for year-to-year growth based on data from national norming studies from six standardized tests in math and science. Their table of average annual gains in effect size (p. 305) served as a benchmark for interpreting the meaning of the effect size in the difference in 11th grade science scores between students in a traditional science sequence and students in a Physics First sequence.

To compare 11th grade science scores of students in a traditional science sequence and students in a Physics First sequence, multiple linear regression was used. Multiple linear regression is appropriate for predicting outcomes when the independent variable is continuous and multiple predictor variables are either categorical or continuous (Field, 2016). Assumptions of linear regression include linearity, independent errors, homoscedasticity, normally distributed errors, and lack of multicollinearity (Field, 2016). The assumption of independent errors was tested using the Durbin-Watson test (Field, 2016). The assumptions of homoscedasticity (the residuals at each level of the predictor variables having the same variance) and linearity were assessed by examining the graphs of standardized residuals and standardized predicted values of the independent variable (Field, 2016). The Variance Inflation Factor (VIF) was used to assess the assumption of non-multicollinearity (Field, 2016). Because large sample sizes are available, the

assumption of normally distributed errors was met (Field, 2016). The first independent variables in the models were the categorical demographic variables for subgroups of historically underserved students (e.g., race/ethnicity, economically disadvantaged status, and English language learner status). Because prior achievement is the strongest predictor of current achievement (Sawyer, 2013), the 8th grade EXPLORE science score was entered as the next predictor in the model. The variable indicating membership in a traditional science or Physics First cohort was then added to the model, followed by interaction terms. The proportion of variance explained by the model, R^2 , was reported to indicate how well each dependent variable in the model predicts the outcome variable, in this case, a student's score on the ACT science test in grade 11 (Field, 2016). The F statistic was used to assess if each model was a significant fit of the data overall (Field, 2016). The t-statistic was used to determine if a predictor made a significant contribution to the outcome using a significance level of $\alpha = .05$ (Field, 2016). The standardized beta values quantified the effect of each predictor variable in the final model on a student's science score in grade 11 (Field, 2016).

Differences between the two groups in student interest in STEM (i.e., students with expressed or implied interest in STEM careers) and in science college- and career-readiness in grade 11 were assessed using binary logistic regression. Binary logistic regression can accommodate multiple predictor variables that can be either categorical or continuous (Field, 2016). Binary logistic regression is appropriate for predicting outcomes when the independent variable is dichotomous and group sizes are unequal (Anderson & Rutkowski, 2008). Assumptions of binary logistic regression are that a

linear relationship exists between any continuous predictors and the logit of the outcome variable and independence of errors (Field, 2016). The first assumption does not apply in this case since the predictor variables are all categorical. Independence of errors was assessed by computing the chi-squared goodness of fit statistic and comparing it to the degrees of freedom (Field, 2016). In addition, contingency tables were run to ensure complete information was available (expected frequencies in each cell are all greater than 1 and less than 20% of cells have frequencies less than 5) (Field, 2016). When more than 20% of cells have frequencies less than 5, Fisher's exact test was used rather than the chi-square test (Field, 2016). The first independent variables in the models were the categorical demographic variables for subgroups of historically underserved students (e.g., race/ethnicity, economically disadvantaged status, and English language learner status). Because prior achievement or interest is the strongest predictor of current interest, 8th grade status was used as the next predictor in the model. The variable indicating membership in a traditional science or Physics First cohort was then added to the model, followed by interaction terms. The Wald statistic was used to determine if a predictor made a significant contribution to the outcome using a significance level of $\alpha = .05$ (Field, 2016). Odds ratios and the 95% confidence interval of the odds ratios quantified how membership in a group affects STEM interest and science college- and career-readiness in grade 11 after controlling for the student's status on each of these measures in 8th grade (Durlak, 2009; Fields, 2016). R^2 was reported using the Cox and Snell's statistic and the Nagelkerke statistic to indicate how well each model predicted the 11th grade outcome (Field, 2016). A summary of the statistical method and predictor

variables for each part of the research question is shown in Figure 4.

ACT Science Score	<ul style="list-style-type: none"> • Method: Multiple Linear Regression • Predictors: Demographics, 8th grade science score, science sequence
Science College Readiness Status	<ul style="list-style-type: none"> • Method: Binary Logistic Regression • Predictors: Demographics, 8th grade science college readiness status, science sequence
STEM Interest	<ul style="list-style-type: none"> • Method: Binary Logistic Regression • Predictors: Demographics, 8th grade STEM interest, science sequence

Figure 4. Research question outcome variables, methods, and predictors.

Conclusion

This retrospective observational study sought to determine if, compared to their peers in a traditional science sequence, historically underserved students completing a Physics First science sequence have higher science test scores, higher rates of science college- and career-readiness, and greater interest in STEM in grade 11. Quasi-experimental designs done well can provide evidence for policy decisions when a true experimental design is not feasible (Campbell & Stanley, 2015; Krathwohl, 2009). Compared to a randomized experiment, a major weakness of quasi-experimental designs is the difficulty in eliminating rival explanations (Campbell & Stanley, 2015; Krathwohl, 2009). To eliminate rival explanations, I conducted a retrospective cohort observational study (Hoffmann & Lim, 2007; Mann, 2003) with a control group and a treatment group consisting of two cohorts each comprised of naturally-formed groups of students based on year of enrollment in 8th grade. In addition, a number of threats to internal validity were effectively controlled using pre-tests measures to confirm similarity of these two

groups (Campbell & Stanley, 2015). Additional strengths of the design included the relatively large number of students in each cohort (more than 2,000) and the use of repeated and vertically scaled measures. Nonetheless, a major confounding variable (Krathwohl, 2009) that could not be controlled is the variability in science instruction students received prior to entering the high school after the November administration of the 8th grade EXPLORE test and the ACT Interest Inventory. Further, some would argue that the assessment of science college- and career-readiness using a single measure like the ACT science test is a limited measure of this construct. Even so, the assessment of the ability of students to reason in science and the quantitative research linking the science college-readiness benchmarks to success in first year science courses cannot be dismissed. Finally, as with the study by Gaubatz (2013), findings from a study conducted in a single school district “should be tempered with the understanding that successful change within educational settings is context-dependent” (p. 25). Nonetheless, I hope districts and schools seeking to close opportunity to learn gaps in science and increase the STEM preparedness of historically underserved students find the results of this study useful due to the methodological strengths of the study, the rigor of the statistical analysis methods, and the disaggregation of results for historically underserved student groups.

CHAPTER 4 RESULTS/ANALYSIS

While there are many factors contributing to disparate educational outcomes for low-income and Black and Hispanic/Latino students, the purpose of this study was to determine if the science college- and career-readiness and interest in STEM of historically underserved students can be improved by implementing a district-wide, inquiry-based high school science sequence comprised of coursework in physics, chemistry, and biology. Historically underserved students are Black, Hispanic/Latino, Native American, Pacific Islander, English language learners, or economically disadvantaged students. This Physics First approach to high school science seeks to close gaps in science content coverage experienced by historically underserved students by ensuring all students are exposed to the same content in physics, chemistry, and biology. As discussed in Chapter 2, incorporating inquiry and modeling develops the academic skills of all students through active learning and the construction of new knowledge; thereby closing gaps in content exposure and emphasis more frequently experienced by historically underserved students compared to their White, economically advantaged peers whose first language is English.

A quantitative retrospective cohort observational study (Hoffmann & Lim, 2007; Mann, 2003) was used to address the primary research question: do historically underserved students completing a Physics First science sequence have 1) higher science college-readiness test scores, 2) higher rates of science college and career readiness, and 3) greater interest in STEM careers in grade 11 compared to their peers who experienced a traditional science sequence? To compare science test scores of students experiencing a

traditional science sequence and students experiencing a Physics First science sequence, multiple linear regression was used. Differences in student interest in STEM (i.e., students with expressed or implied interest in STEM careers as described in Chapter 3) and science college- and career-readiness of students experiencing a traditional science sequence and students experiencing a Physics First science sequence were assessed using binary logistic regression. A number of threats to internal validity were controlled through the inclusion of pre-measures from 8th grade in the regression models to control for any initial differences between the two groups (Campbell & Stanley, 2015). Additional strengths of the design included the large number of students (more than 4,500) in each group and the use of repeated and vertically scaled measures of science college-readiness. I now present the results of the data analysis and the statistical methods used to derive the results.

Analysis of Data and Presentation of Results

A major disadvantage of a retrospective cohort observational study is the inability to control for all factors that may differ between the two groups (Hoffmann & Lim, 2007; Mann, 2003). Hoffmann and Lim (2007), Mann (2003) and others refer to these factors as confounding variables. Multivariate models, including linear and logistic regression models, can be used to control for such confounding variables (Hoffmann & Lim, 2007; Pourhoseingholi, Baghestani, & Vahedi, 2012). An additional potential problem in this study was bias. Mann (2003) asserts that “bias can occur in any research and reflects the potential that the sample studied is not representative of the population it was drawn from and/or the population at large” (p. 55). To identify sources of bias, I begin the data

analysis with a thorough examination of the demographic differences between the students who began 9th grade in a traditional science sequence and those students who entered a Physics First sequence, including the attrition of students over the course of their first three years of high school. I then examine initial differences of the two groups on the 8th grade pre-measures of science test score, science college- and career-readiness status, and STEM interest. I conclude the analysis by presenting the results of the regression models that control for any initial differences.

Differences in 9th Grade Demographics. Enrollment in grade 9 on May 1 is shown in Table 3. Students entering 9th grade in the fall of 2010 and 2011 experienced a

Table 3

Demographics of 9th Grade Students by Cohort

		Traditional Science		Physics First		
Student Group		Fall 2010	Fall 2011	Fall 2012	Fall 2013	Total
Native American	Count	20 _a	14 _a	16 _a	18 _a	68
	% within Group	0.6%	0.5%	0.6%	0.6%	0.6%
Asian	Count	377 _a	381 _a	374 _a	378 _a	1510
	% within Group	12.1%	13.1%	12.9%	13.0%	12.8%
Black	Count	86 _a	87 _a	80 _a	80 _a	333
	% within Group	2.8%	3.0%	2.8%	2.7%	2.8%
Latino	Count	690 _a	629 _a	648 _a	683 _a	2650
	% within Group	22.1%	21.6%	22.3%	23.5%	22.4%
Pacific Islander	Count	28 _a	28 _a	23 _a	22 _a	101
	% within Group	0.9%	1.0%	0.8%	0.8%	0.9%
Multiracial	Count	191 _a	182 _a	215 _a	215 _a	803
	% within Group	6.1%	6.3%	7.4%	7.4%	6.8%
Economically Disadv.	Count	1288 _a	1117 _b	1165 _{a, b}	1183 _{a, b}	4753
	% within Group	41.3%	38.4%	40.1%	40.7%	40.1%
English Lang. Learner	Count	292 _a	196 _b	178 _b	174 _b	840
	% within Group	9.4%	6.7%	6.1%	6.0%	7.1%

Each subscript letter denotes a subset of Student Group categories whose column proportions do not differ significantly from each other at the .05 level.

traditional science sequence. Students entering 9th grade in the fall of 2012 and 2013 experienced a Physics First curriculum with modeling. Table 4 displays the percentage of students in each cohort enrolled for three consecutive years in grades 9, 10, and 11.

Table 4

Demographics of 9th Grade Students Enrolled Three Years by Cohort

		Traditional Science		Physics First		Total
		Fall 2010	Fall 2011	Fall 2012	Fall 2013	
Native Am.	Count	14 _a	10 _a	10 _a	14 _a	48
	% within Group	0.6%	0.4%	0.4%	0.6%	0.5%
Asian	Count	324 _a	345 _a	341 _a	335 _a	1345
	% within Group	13.8%	14.6%	14.1%	13.7%	14.0%
Black	Count	55 _a	63 _a	53 _a	63 _a	234
	% within Group	2.3%	2.7%	2.2%	2.6%	2.4%
Latino	Count	465 _a	466 _a	510 _{a, b}	551 _b	1992
	% within Group	19.8%	19.7%	21.0%	22.6%	20.8%
Pacific Islander	Count	15 _a	13 _a	19 _a	17 _a	64
	% within Group	0.6%	0.5%	0.8%	0.7%	0.7%
Multiracial	Count	147 _a	159 _a	177 _a	187 _a	670
	% within Group	6.3%	6.7%	7.3%	7.7%	7.0%
Economically Disadv.	Count	805 _{a, b}	785 _b	864 _{a, b}	898 _a	3352
	% within Group	34.4%	33.2%	35.6%	36.8%	35.0%
English Lang. Learner	Count	168 _a	135 _b	115 _b	130 _b	548
	% within Group	7.2%	5.7%	4.7%	5.3%	5.7%

Each subscript letter denotes a subset of Group categories whose column proportions do not differ significantly from each other at the .05 level.

The results in Table 5 suggest that the populations of 9th grade students enrolled on May 1 for three consecutive years in each cohort are significantly different ($\alpha=.05$) for all economically disadvantaged students ($p<.05$) and English language learners ($p<.01$). Table 4 reveals that the percentage of English language learners enrolled in the first cohort (7.2%) is significantly higher ($\alpha=.05$) compared to the other three cohorts. Table 4

also shows that the percentages of economically disadvantaged, and Hispanic/Latino students are significantly different between some of the cohorts.

Table 5

Pearson Chi-Square Tests on Cohort Demographics by Enrollment Status

	All Students	Enrolled Three Years
Student Group	Value	Value
Native American (AI)	0.816	1.317
Asian (AS)	1.708	0.839
Black (BL)	0.457	1.429
Hispanic/Latino (Latino)	3.057	7.835
Pacific Islander (PI)	0.947	1.050
Multiracial (MU)	6.784	4.166
Economically Disadvantaged (ECD)	5.755	7.839 *
English Language Learners (ELL)	34.676 ***	14.138 **

df = 3 *p< .05 **p<.01 ***p< .001

Enrollment in grade 9 on May 1 and enrollment on May 1 for three consecutive years by science sequence is shown in Table 6. Students in the 2010 and 2011 cohorts are labeled traditional science and students in the cohorts of 2012 and 2013 are labeled Physics First. Pearson chi-square tests of significance for students enrolled on May 1 are shown in Tables 7. The results in Table 7 again suggest that the populations of students enrolled on May 1 in grade 9 are significantly different ($\alpha=.05$) only for English language learners. The results in Table 7 suggest that the populations of 9th grade students enrolled on May 1 for three consecutive years are significantly different ($\alpha=.05$) for the Hispanic/Latino, economically disadvantaged, and English language learner student groups. The percentage of students enrolled for three years who are English language learners are higher in the traditional science group while the percentages of students who are economically disadvantaged or Hispanic/Latino are higher in the Physics First group.

By including demographic variables in the regression models, these differences were controlled for (McNamee, 2005).

Table 6
9th Grade Demographics by Science Sequence

Student Group	All students				Enrolled Three Years			
	Traditional Science		Physics First		Traditional Science		Physics First	
	n	%	n	%	n	%	n	%
Native American/Alaskan Native	34	0.6	34	0.6	24	0.5	24	0.5
Asian	758	12.6	752	12.9	669	14.2	676	13.9
Black	173	2.9	160	2.8	118	2.5	116	2.4
Hispanic/Latino	1319	21.9	1331	22.9	931	19.8	1061	21.8
Pacific Islander/Native Hawaiian	56	0.9	45	0.8	28	0.6	36	0.7
White	3312	55.0	3066	52.7	2635	55.9	2589	53.2
Multiracial	373	6.2	430	7.4	306	6.5	364	7.5
Economically Disadvantaged	2405	39.9	2348	40.4	1590	33.8	1762	36.2
English Language Learner	488	8.1	352	6.1	303	6.4	245	5.0
Historically Underserved	2857	47.4	2775	47.7	1930	41.0	2117	43.5
Total	6025	100.0	5818	100.0	4711	100.0	4866	100.0

Table 7
Pearson Chi-Square Tests of Student Demographics by Science Sequence

Student Group	All Students	Students Enrolled Three Years
	Value	Value
Native American (AI)	0.021	0.013
Asian (AS)	0.316	0.189
Black (BL)	0.159	0.147
Hispanic/Latino (Latino)	1.654	6.059 *
Pacific Islander (PI)	0.852	0.763
Multiracial (MU)	6.743 **	3.570
Economically Disadvantaged (ECD)	0.239	6.365 *
English Language Learners (ELL)	18.865 ***	8.657 **
df = 1	*p< .05	**p<.01 ***p<.001

Differences in demographics due to attrition. As may be expected, the percentages of economically disadvantaged students enrolled for three consecutive years

on May 1 was 4 to 6 percentage points lower than the total 9th grade enrollment shown in Table 6. Students who are economically disadvantaged are less likely to attend school in the same district due to the economic challenges facing lower income families (Temple & Reynolds, 2000; Voight, Shinn, & Nation, 2012). Z-scores for each population reported in Table 8 suggest that the differences in the percentages of economically disadvantaged, English language learners, and Hispanic/Latino students in the traditional science and Physics First groups enrolled for three years are not equivalent to the percentages of 9th grade students enrolled on May 1, with the exception of Hispanic/Latino students in the Physics First group. While these differences are statistically significant ($\alpha=.05$) and may represent sample bias, in order to answer the research questions for this study, students must be enrolled for three consecutive years in order to identify and quantify any differences in student outcomes between the two science sequences.

Table 8
Z-Scores of Demographics of Students by Enrollment Status

	Traditional Science	Physics First
Student Group	Z-score	Z-score
Native American (AI)	0.3849	0.6387
Asian (AS)	-2.4535 **	-1.4627
Black (BL)	1.1608	1.1883
Hispanic/Latino (Latino)	2.6907 **	1.3249
Pacific Islander (PI)	1.9556	0.1996
Multiracial (MU)	-0.6434	-0.1759
Economically Disadvantaged (ECD)	4.3878 ***	6.5993 ***
English Language Learners (ELL)	3.2826 **	2.2753 **

*p< .05 **p<.01 ***p<.001

Differences in 8th grade science test scores, college-readiness, and STEM

interest. Science test scores, science college-readiness, and STEM interest in the 8th grade year were used to control for initial group differences (Tuckman, 1994). Because

the standard error of measurement is approximately 2 scale score points for each of the subject-area test scores (ACT, 2014), differences of more than 1 scale score point on the 8th grade scores would suggest non-equivalent groups. As shown in Table 9, the mean 8th grade science scores of students entering a traditional science sequence and students entering a Physics First science sequence are within 0.3 points, less than one-tenth of a standard deviation of the mean scores in either group. Mean 8th grade science scores for students who are Native American, Black, Hispanic/Latino, Pacific Islander, economically disadvantaged, and English language learners in the traditional science instruction cohorts and students in the Physics First cohorts are all within 0.5 points of each other and are less than two-tenths of a standard deviation. With the exception of Native American students, the differences in 8th grade mean scale scores are higher for students receiving traditional science instruction.

Independent sample t-tests were performed on the mean 8th grade science scores for each of the historically underserved student groups to determine if there were statistically significant differences in the mean 8th grade science scores between students in the traditional science cohorts and students in the Physics First cohorts. A two-tailed t-test was employed with a significance level of $\alpha = .05$. Differences in mean 8th grade science scores between the traditional science cohorts and the Physics First cohorts were statistically significant for all students, $t(8367) = 2.925$, $p = .003$, and for economically disadvantaged students, $t(2804) = 2.871$, $p = .004$. Differences in mean 8th grade science scores between the traditional science cohorts and the Physics First cohorts were not statistically significant for Native American students, $t(37) = -0.479$, $p = .635$; Black

Table 9
Mean Science Scores by Year

Grade	Traditional Science			Physics First		
	N	M	SD	N	M	SD
All Students						
8	4089	18.11	3.444	4286	17.89	3.704
9	4333	19.30	3.662	4339	19.27	3.707
10	4221	20.80	4.588	4203	20.85	4.885
11	4229	21.55	5.744	4349	21.91	5.937
Native American						
8	19	16.74	3.478	20	17.25	3.210
9	21	19.29	3.379	23	18.39	3.738
10	20	20.10	3.919	22	19.32	4.099
11	23	20.83	5.852	22	20.77	4.710
Black						
8	83	15.82	2.812	89	15.53	3.425
9	103	16.63	3.178	95	16.80	3.512
10	99	17.17	3.273	89	17.60	3.878
11	99	17.07	4.415	95	17.79	5.329
Hispanic/Latino						
8	786	15.95	2.948	939	15.71	3.249
9	805	16.99	3.376	916	16.99	3.367
10	755	17.92	3.708	868	17.82	3.898
11	764	17.55	5.038	911	17.99	4.967
Pacific Islander						
8	23	15.96	2.513	33	15.94	3.544
9	25	17.28	4.005	32	17.50	4.197
10	23	18.48	4.621	31	18.19	5.782
11	25	18.72	4.852	29	18.48	6.294
Economically Disadvantaged						
8	1307	16.37	3.066	1515	16.02	3.296
9	1379	17.42	3.509	1502	17.30	3.485
10	1306	18.46	3.923	1420	18.28	4.162
11	1307	18.44	5.220	1474	18.50	5.141
English Language Learner						
8	183	13.75	2.271	139	13.25	2.774
9	225	14.74	3.190	165	14.36	2.988
10	223	15.82	3.001	169	15.18	2.904
11	232	14.61	4.206	193	15.18	4.420

students, $t(170) = 0.607$, $p = .545$; Hispanic/Latino students, $t(1723) = 1.612$, $p = .107$; Pacific Islander students, $t(54) = 0.020$, $p = .984$; and English language learners, $t(320) = 1.766$, $p = .078$. To control for initial differences, 8th grade science scores were included in the multiple linear regression model.

The percentage of students who were college-ready in science in 8th grade, shown in Table 10, differed by 4 percentage points, with a lower percentage of students in the Physics First sequence meeting the science college readiness benchmark (49.6%) than students in the traditional science sequence (53.6%). Differences in science college readiness between students in the traditional science cohorts and students in the Physics First cohorts were also within 5 percentage points for the Native American, Black, Hispanic/Latino, economically disadvantaged, and English language learners student groups. As with the mean science scores, the percentage of students who met the college-readiness benchmark in science in grade 8 were higher for students in the traditional science cohorts compared with students in the Physics First cohorts for all of these students groups, except Native Americans. For Pacific Islanders, there was a 12 percentage point difference in favor of students entering the traditional science sequence. Chi-square tests (Table 11) suggest that these differences in the percentages of students scoring at or above the college readiness benchmark of 18 on the EXPLORE science test in grade 8 in the traditional science cohorts and in the Physics First cohorts were not equivalent ($\alpha = .05$) for the all students ($p < .001$) and economically disadvantaged groups ($p < .01$). The percentages of science college readiness of Native American, Black, and Hispanic/Latino students in the traditional science cohorts and in the Physics First cohorts

were not significantly different at $\alpha = .05$. The percentages of science college readiness of Pacific Islanders and English language learners in the traditional science cohorts and in the Physics First cohorts were also not significantly different at $\alpha = .05$ using Fisher's exact test. Fisher's exact test computes the exact probability of the chi-square statistic when one or more cells have frequencies less than 5 (Field, 2016).

Table 10

College Ready in Science by Year

Grade	All Students		Native American		Black		Hispanic/Latino		Pacific Islander		Economically Disadvantaged		English Language Learners	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
<i>Traditional Science</i>														
8	2191	53.6	7	36.8	20	24.1	190	24.2	6	26.1	407	31.1	4	2.2
9	2545	58.7	13	61.9	31	30.1	242	30.1	9	36.0	516	37.4	20	8.9
10	2467	58.4	10	50.0	24	24.2	217	28.7	8	34.8	469	35.9	29	13.0
11	1893	44.8	10	43.5	15	15.2	128	16.8	6	24.0	296	22.6	12	5.2
<i>Physics First</i>														
8	2126	49.6	8	40.0	19	21.3	208	22.2	5	15.2	404	26.7	2	1.4
9	2474	57.0	13	56.5	23	24.2	267	29.1	10	31.3	491	32.7	6	3.6
10	2392	56.9	7	31.8	24	27.0	240	27.6	9	29.0	460	32.4	11	6.5
11	1997	45.9	6	27.3	15	15.8	159	17.5	7	24.1	327	22.2	11	5.7

Table 11

Pearson Chi-Square Tests on 8th Grade Science College Readiness

Student Group	N	Value	Significance
All students	8375	13.268	.000 ***
Native American	39	5.449	.839
Black	172	0.185	.667
Hispanic/Latino	1725	0.985	.321
Pacific Islander	56		.249 ^a
Economically Disadvantaged	2822	6.856	.009 **
English Language Learners	322		.702 ^b

df=1 *p<.05 **p<.01 ***p<.001

a. 1 cell (25.0%) had expected count less than 5; Fisher's exact statistic reported.

b. 2 cells (50.0%) had expected count less than 5; Fisher's exact statistic reported.

The percentages of students interested in STEM (expressed or measured) in 8th grade for students in the traditional science sequence and in the Physics First science sequence are within 5 percentage points for all students, Hispanic/Latino students, economically disadvantaged students, and English language learners (Table 12). For the Native American and Black student groups, the differences are less than 10 percentage points. With the exception of Pacific Islander students, where the difference was 13 percentage points, a greater percentage of students were interested in STEM prior to entering the Physics First sequence compared to the students entering the traditional science sequence. The results of chi-square tests shown in Table 13 are all non-significant at $\alpha = .05$ and thus do not detect statistically significant differences between the groups in each case. Nonetheless, STEM interest in 8th grade was included in the

Table 12
STEM Interest by Year

Grade	All Students		Native American		Black		Hispanic/Latino		Pacific Islander		Economically Disadvantaged		English Language Learners	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
<i>Traditional Science</i>														
8	2037	52.4	7	38.9	40	51.3	381	52.2	8	44.4	636	52.2	636	52.2
9	2112	51.8	9	47.4	49	52.7	359	48.6	8	44.4	629	50.2	629	50.2
10	2028	50.6	8	42.1	44	50.0	314	45.7	10	52.6	576	48.0	576	48.0
11	1423	45.3	7	41.2	27	40.9	211	40.0	6	50.0	393	43.2	393	43.2
<i>Physics First</i>														
8	2191	53.8	9	47.4	47	57.3	469	53.5	10	31.3	750	52.9	70	55.1
9	2136	54.4	8	40.0	42	48.8	413	51.2	8	30.8	689	51.9	60	44.4
10	2127	54.0	7	36.8	38	46.3	398	50.6	14	56.0	685	52.6	67	48.2
11	1897	52.5	5	35.7	32	44.4	332	47.2	14	66.7	559	49.3	65	50.8

Table 13

Pearson Chi-Square Tests on 8th Grade STEM Interest

Student Group	N	Value
All students	7962	1.481
Native American	37	0.271
Black	160	0.587
Hispanic/Latino	1606	0.290
Pacific Islander	50	0.870
Economically Disadvantaged	2637	0.135
English Language Learners	291	0.583

df=1

regression model to control for the non-significant differences between the groups. As Tryon (2001) correctly notes “absence of positive evidence for statistical difference does not constitute presence of positive evidence for statistical equivalence” (p. 379).

Results of regressions and analysis of outcomes. To control for initial differences in demographics, science achievement and STEM interest between students entering a traditional science sequence and students entering a Physics First sequence with modeling, regression models developed for this study included these potential confounding variables. I now describe the development of the models and the results for each of the three measures of interest: ACT science scores, science college- and career-readiness status, and STEM interest in grade 11.

Mean 11th grade science scores. Multiple linear regression was used to answer the first part of the research question -- do historically underserved students in a Physics First science sequence have higher ACT science scores in grade 11 than their peers in a traditional science sequence. Multiple linear regression is appropriate for predicting outcomes when the dependent variable is continuous and multiple predictor variables are

either categorical or continuous (Field, 2016). Student characteristics (i.e., race/ethnicity, economic disadvantaged status, and English language learner status) were entered into the model first. Because prior achievement is the strongest predictor of current achievement (Sawyer, 2013), the 8th grade EXPLORE science score was entered as the next predictor in the model. The variable capturing whether a student entered 9th grade in the traditional science sequence or in the Physics First sequence was entered as the next predictor. Finally, two-way interaction terms were entered into the model. Three-way actions were entered for any two-way interaction terms that were significant. The proportion of variance explained by the model, R^2 , explains how well each independent variable in the model predicts a student's score on the ACT science test in grade 11 (Field, 2016). The F statistic was used to assess if each model was a significant fit of the data overall (Field, 2016). The t-statistic indicated if a predictor makes a significant contribution to the outcome using a significance level of $\alpha = .05$ (Field, 2016). Effect sizes for the differences in the science mean scores are reported using the standardized beta coefficients.

The linear regression model (see Appendix B) predicts that the 11th grade mean science score for a student in a Physics First sequence to be .74 points higher than for a student in a traditional science sequence, regardless of the student's 8th grade science score ($p < .001$). The 95% confidence intervals of the estimate are .53 and .95. Adding the treatment group status increased R^2 by .003 to .561 and was a significant improvement in the model (change in $F = 51.783$, $p < .001$). However, economically disadvantaged students in the Physics First sequence benefitted less than other student

groups. The interaction term of enrollment in Physics First and being economically disadvantaged was $-.45$ and was also significant ($p = .016$). Thus, for economically disadvantaged students, the difference in 11th grade science score for students in Physics First was only $.29$ ($.74 - .45$) points higher than for their economically disadvantaged peers in a traditional science sequence.

Interaction terms of a student's demographic and the student's 8th grade science score was significant at $\alpha=.05$ for Black, Hispanic/Latino, economic disadvantaged, and English language learner students groups, indicating in each case that students from these groups experience lower 11th grade science scores compared to the reference group of White students with the same initial 8th grade score. Adding the interaction term of Physics First enrollment and 8th grade science score did not significantly improve the model at $\alpha=.05$ ($t=-0.242$, $p = .808$), which suggests that Physics First benefits all students regardless of a student's 8th grade science score. The full final regression model is reported in Appendix B.

Science college- and career-readiness in grade 11. The science college- and career-readiness status of three in four students in both the traditional science sequence and the Physics First sequence did not change between the fall of 8th grade and spring of 11th grade. In both the traditional science sequence and the Physics First sequence, across all populations with more than 10 students meeting the science college readiness benchmark in grade 8, greater percentages of students moved from college ready to not college ready between 8th and 11th grade than the reverse. With the exception of the English language learners group (which is comprised of very few students meeting

science college-readiness benchmarks at either grade 8 or grade 11), the differences in the percentages of students losing science college readiness status versus students gaining science college-readiness status were narrower for students experiencing a Physics First science sequence with modeling. For example, 15.5% of economically disadvantaged students in the traditional science sequence met the science college-readiness benchmark in grade 8 but failed to do so in grade 11 compared to the 5.8% who were not science college-ready in grade 8 but met the science college-readiness benchmark in grade 11 (see Table 14), a difference of 9.7 percentage points. In contrast, for economically disadvantaged students in the Physics First sequence with modeling, the corresponding difference was 5.6 percentage points (10.6% - 5.0%).

To answer the second part of the research question -- do historically underserved students completing a Physics First science sequence have higher rates of science college- and career-readiness compared to their peers who experienced a traditional science sequence -- differences in science college- and career-readiness in 11th grade between the groups were assessed using binary logistic regression. Binary logistic regression is appropriate for predicting outcomes when the dependent variable is dichotomous and group sizes are unequal (Anderson & Rutkowski, 2008). Binary logistic regression can accommodate multiple predictor variables that can be either categorical or continuous (Field, 2016). Student characteristics (i.e., race/ethnicity, economic disadvantaged status, and English language learner status) were entered first into the model as predictors. Because prior achievement is the strongest predictor of current achievement (Sawyer, 2013), 8th grade science college readiness status was used

Table 14

Science College Readiness Status in Grades 8 and 11 by Science Sequence

College Ready in Science Grade 11		Traditional Science		Physics First	
		College Ready in Science Grade 8			
		No	Yes	No	Yes
	All Students				
	No	1442	611	1605	481
		38.2%	16.2%	41.0%	12.3%
	Yes	250	1467	310	1522
		6.6%	38.9%	7.9%	38.8%
	Native American				
	No	10	1	10	6
	55.6%	5.6%	50.0%	30.0%	
Yes	1	6	2	2	
	5.6%	33.3%	10.0%	10.0%	
Black					
No	49	12	55	6	
	69.0%	16.9%	75.3%	8.2%	
Yes	3	7	2	10	
	4.2%	9.9%	2.7%	13.7%	
Hispanic/Latino					
No	466	95	600	86	
	68.6%	14.0%	72.4%	10.4%	
Yes	36	82	35	108	
	5.3%	12.1%	4.2%	13.0%	
Pacific Islander					
No	16	1	19	0	
	72.7%	4.5%	73.1%	0.0%	
Yes	0	5	2	5	
	0.0%	22.7%	7.7%	19.2%	
Economically Disadvantaged					
No	690	176	875	140	
	60.9%	15.5%	66.5%	10.6%	
Yes	66	201	66	235	
	5.8%	17.7%	5.0%	17.9%	
English Language Learners					
No	145	3	121	1	
	95.4%	2.0%	96.8%	0.8%	
Yes	3	1	2	1	
	2.0%	0.7%	1.6%	0.8%	

College Ready in Science Grade 11

as the next predictor in the model. The variable capturing whether a student entered 9th grade in the traditional science sequence or in the Physics First sequence was entered as the next predictor. Finally, two-way interaction terms were entered into the model. Three-way actions were entered for any two-way interaction terms that were significant. Using a significance level of $\alpha = .05$, the Wald statistic indicated if a predictor made a significant contribution to the model for 11th grade science college readiness (Field, 2016). Odds ratios and the 95% confidence interval of the odds ratios are reported to quantify how membership in a group influences science college- and career-readiness after controlling for the student's science college- and career-readiness status in grade 8 (Durlak, 2009; Fields, 2016). R^2 is reported using the Cox and Snell's statistic and the Nagelkerke statistic to indicate how well each model predicts the outcome (Field, 2016).

Students in the Physics First sequence were 1.28 times as likely to be science college ready in grade 11 as students in the traditional science sequence ($\chi^2(1)=18.146$, $p<.001$). The 95% confidence interval of this odds ratio is 1.14 and 1.43. Including the interaction of science college readiness in grade 8 and membership in the Physics First sequence in the model was not significant at $\alpha=.05$ ($\chi^2(1)=1.544$, $p = .214$).

Interaction terms of a student's demographic and the student's 8th grade college readiness status were not significant ($\alpha=.05$) for any of the groups of students who are traditionally underserved in STEM. This suggests that students in Physics First are more likely to be college and career ready in science in grade 11 compared to their peers in a traditional science sequence regardless of their demographics. Adding the interaction term of Physics First enrollment and 8th grade science college readiness status did not

significantly improve the model which implies that Physics First with modeling benefits students regardless of a student's 8th grade science college- and career-readiness status.

The final logistic regression model is reported in Appendix C.

STEM interest in grade 11. For 65% of students with career choices and interest inventory scores, interest in STEM, whether expressed or measured, did not change between the fall of 8th grade and spring of 11th grade in both the control and treatment groups. Across all populations in both the traditional science sequence and the Physics First sequence, greater percentages of students lost interest in STEM careers between 8th and 11th grade than gained interest. Across all groups, the differences in the percentages of students losing interest versus students gaining interest in STEM were narrower for students experiencing a Physics First science sequence with modeling. For example, 23.5% of economically disadvantaged students in the traditional science sequence lost interest in STEM compared to 13.9% who gained interest between grades 8 and 11 (see Table 15), a difference of 9.6 percentage points. In contrast, for economically disadvantaged students in the Physics First with modeling sequence, the corresponding difference is 4.7 percentage points (21.2% - 16.5%).

Differences in STEM interest in 11th grade between students in the traditional science sequence and students in the Physics First sequence with modeling were assessed using binary logistic regression in order to answer the third part of the research question - do historically underserved students completing a Physics First science sequence have greater interest in STEM careers compared to their peers who experienced a traditional science sequence? Binary logistic regression is appropriate for predicting outcomes

Table 15

STEM Interest in Grade 8 and 11 by Science Sequence

STEM Interest in Grade 11		Traditional Science		Physics First	
		STEM Interest in Grade 8			
		No	Yes	No	Yes
		All Students			
	No	894	571	897	555
		32.8%	20.9%	28.8%	17.8%
	Yes	375	887	515	1153
		13.8%	32.5%	16.5%	37.0%
		Native American			
	No	4	2	4	4
	33.3%	16.7%	33.3%	33.3%	
Yes	3	3	1	3	
	25.0%	25.0%	8.3%	25.0%	
	Black				
No	18	14	17	12	
	37.5%	29.2%	31.5%	22.2%	
Yes	5	11	8	17	
	10.4%	22.9%	14.8%	31.5%	
	Hispanic/Latino				
No	156	116	177	146	
	34.6%	25.7%	28.8%	23.8%	
Yes	65	114	98	193	
	14.4%	25.3%	16.0%	31.4%	
	Pacific Islander				
No	2	3	4	3	
	18.2%	27.3%	21.1%	15.8%	
Yes	2	4	8	4	
	18.2%	36.4%	42.1%	21.1%	
	Economically Disadvantaged				
No	247	179	280	205	
	32.4%	23.5%	28.9%	21.2%	
Yes	106	230	160	323	
	13.9%	30.2%	16.5%	33.4%	
	English Language Learners				
No	28	20	21	17	
	32.6%	23.3%	29.6%	23.9%	
Yes	13	25	13	20	
	15.1%	29.1%	18.3%	28.2%	

when the dependent variable is dichotomous and group sizes are unequal (Anderson & Rutkowski, 2008). Binary logistic regression can accommodate multiple predictor variables that can be either categorical or continuous (Field, 2016). Student characteristics (i.e., race/ethnicity, economic disadvantaged status, and English language learner status) were entered first into the model as predictors. To control for any initial group differences, STEM interest in grade 8 was used as the next predictor in the model. The variable capturing whether a student entered 9th grade in the traditional science sequence or in the Physics First sequence was entered as the next predictor. Finally, two-way interaction terms were entered into the model. Three-way actions were entered for any two-way interaction terms that were significant. Using a significance level of $\alpha = .05$, the Wald statistic indicated if a predictor made a significant contribution to model for 11th grade STEM interest (Field, 2016). Odds ratios and the 95% confidence interval of the odds ratios are reported to quantify how membership in a group influences STEM interest in grade 11 after controlling for the student's STEM interest in grade 8 (Durlak, 2009; Fields, 2016). R^2 is reported using the Cox and Snell statistic and the Nagelkerke statistic to indicate how well each model predicts the outcome (Field, 2016).

Students in Physics First were 1.37 times as likely to be interested in STEM in grade 11 as students in a traditional science sequence ($\chi^2(1)=29.694$, $p<.001$). The 95% confidence interval of this odds ratio is 1.23 and 1.53. The interaction of STEM interest in grade 8 and enrollment in Physics First was not significant ($\chi^2(1)=0.064$, $p = .800$) which suggests that, compared to a traditional science sequence, Physics First increases student interest in STEM regardless of a student's interest in STEM in grade 8.

Interaction terms of a student's demographic and the student's 8th STEM interest were not significant ($\alpha=.05$) for any of the groups of students who are traditionally underserved in STEM with the exception of Latino students and Pacific Islanders. This suggests that for all other demographic groups, students in Physics First are more likely to be interested in STEM in grade 11 compared to their peers in a traditional science sequence. Latino students interested in STEM in grade 8 are less likely to be interested in STEM in grade 11 compared to their White peers whether they were in a traditional science sequence or a Physics First sequence in high school. Drawing a similar conclusion for Pacific Islander students is more problematic due to the small number of students in the data set. Adding the interaction term of demographics and Physics First enrollment did not significantly improve the model, suggesting that Physics First increases the likelihood a student will be interested in STEM in grade 11 regardless of his/her demographic group membership. The final logistic regression model is reported in Appendix C.

Interpretation of Findings

Do historically underserved students completing a Physics First science sequence have 1) higher science test scores, 2) higher rates of science college- and career-readiness, and 3) greater interest in STEM careers compared to their peers who experienced a traditional science sequence? The results of the multiple linear regression and binary logistic regressions outlined above suggest the answer is a qualified yes.

Science test scores. The 11th grade science score for students experiencing a Physics First with modeling science sequence was 0.74 points higher than that of students

in a traditional science sequence when controlling for any initial differences. Figure 5 puts this difference in context in two ways. First, the science college- and career-readiness benchmark increases by 5 points between grades 8 and 11, so this difference represents 15% of the increased learning characterized by these college-readiness expectations. Second, the mean science gain scores of students in the traditional science sequence are shown in the same figure. The effect of enrollment in Physics First is equivalent to 21% of the mean gain science scores between grades 8 and 11 of students in a traditional science sequence and almost half or more of the mean gain scores for Black and Hispanic/Latino students. A nationally representative sample of approximately 150,000 students reported that the average growth on the ACT science test between grade 8 and grade 12 is 3.3 points (ACT, 2009b).

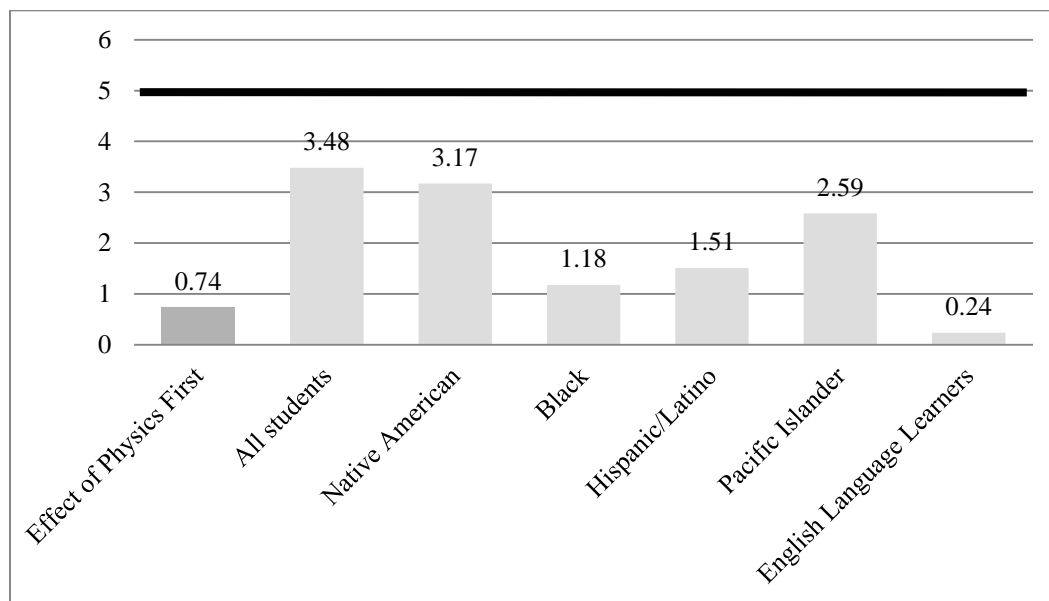


Figure 5. Effect of Physics First on 11th grade ACT science score. 11th grade score difference predicted by the multiple linear regression model between students in a Physics First sequence compared to students in a traditional science sequence (bar at left) compared to the mean gain score of students in traditional science and to the increase in science college readiness benchmark between grades 8 and 11 (black line).

Standardized beta coefficients are used as standardized effect sizes for multiple linear regression (Kelley & Preacher, 2012). The effect size of these differences between students in the Physics First with modeling sequence and students in a traditional science sequence was .06. This effect size represents small positive effects in the context of student achievement gains in high school on standardized tests of science (Bloom et al., 2008). For economically disadvantaged students, however, the 11th grade science score for students experiencing a Physics First with modeling science sequence was only 0.29 points higher than that of students in a traditional science sequence when controlling for initial differences, less than half the effect for all students and for other historically underserved populations. As illustrated in Figure 6, the effect of Physics First is less than 15% of the mean gain science score from grade 8 to grade 11 for economically disadvantaged students in a traditional science sequence.

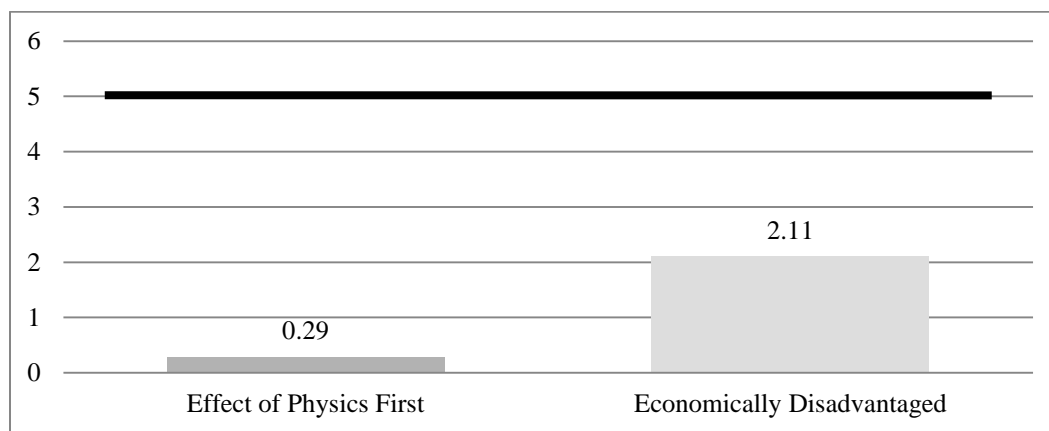


Figure 6. Effect of Physics First for economically disadvantaged students. 11th grade score difference predicted by the multiple linear regression model between economically disadvantaged students in a Physics First sequence and economically disadvantaged students in a traditional science sequence (bar at left) compared to the mean gain score of economically disadvantaged students in traditional science (bar at right) and to the increase in science college readiness benchmark between grades 8 and 11 (black line).

Science college readiness. Students experiencing a Physics First science sequence with modeling, including historically underserved student, were 1.28 times as likely to meet the college- and career-readiness benchmark on the ACT science test as their peers experiencing a traditional science sequence. This odds ratio and its 95% confidence interval are shown on the left in Figure 7.

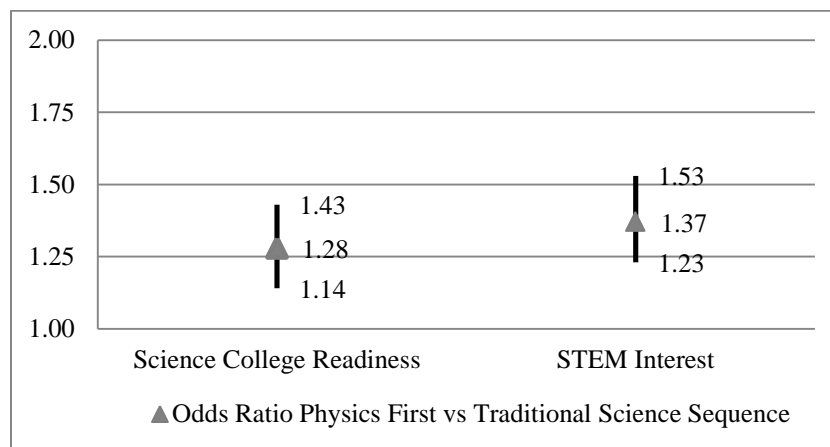


Figure 7. Odds ratios produced by the binary logistic regressions. Odds ratios shown with 95% confidence intervals quantify how many times as likely a student in a Physics First science sequence is to achieve an outcome compared to a student in a traditional science sequence after controlling for initial differences in demographics and 8th grade status.

Interest in STEM. Students experiencing a Physics First with modeling science sequence, including historically underserved student, were 1.37 times as likely to express interest in STEM in grade 11 as their peers experiencing a traditional science sequence. This odds ratio and its 95% confidence interval are shown on the right side in Figure 7.

After controlling for initial differences in 8th grade, this study found historically underserved students in a Physics First science sequence had higher ACT science test scores, higher rates of science college- and career-readiness, and greater interest in STEM careers in grade 11 than their peers in a traditional science sequence. Unfortunately, for

economically disadvantaged students, the 11th grade science score difference between students experiencing a Physics First with modeling science sequence and students in a traditional science sequence was less than half the effect for all students and for other historically underserved populations. However, as with other retrospective cohort observational studies, these results have several limitations (Mann, 2003).

Limitations of the Study

Compared to a randomized experiment, a major weakness of quasi-experimental designs is the difficulty in eliminating rival explanations (Campbell & Stanley, 2015; Krathwohl, 2009). Multiple linear regression and binary logistic regression models were used to account for any initial differences between students in the traditional science cohorts and students in the Physics First cohorts in demographic as well as 8th grade science scores, science college- and career-readiness status, and interest in STEM. Other possible confounding variables could not be controlled, however. These variables, which weaken the internal validity of the study (Krathwohl, 2009), included any differences between the two cohorts in science instruction received in the 8th grade between November and June and changes in teachers assigned to courses and their initial content and pedagogical knowledge. It is important to keep in mind that this district implemented a Physics First sequence in response to the NGSS. Professional development in the NGSS for 8th grade science teachers could influence the instruction that occurred in 8th grade science occurring between November and June. In addition, it is possible that instructional changes in other content areas during the study period such as mathematics or English language arts could influence student achievement in science.

This study also did not adjust for the earlier 11th grade ACT testing experienced by students in the first cohort of the Physics First sequence, which may underestimate the effects of Physics First.

The use of two cohorts in each group and the large number of students in each group provided strong observed statistical power for the regression analysis (Soper, 2016). However, the small numbers of students in the Native American and Pacific Islander groups warrant interpreting findings for these groups with caution.

While the college readiness benchmarks established for the EXPLORE and ACT science tests predict a student's success in credit-bearing college coursework in college biology, the use of a single assessment alone to predict college readiness is "imperfect, incomplete, and limited in what it can assess" (Maruyama, 2012, p. 254). However, single dimension benchmarks have advantages, including, in the case of the ACT science assessment, providing more information about a student's areas of need compared to high school GPA (Mattern, Radunzel, & Westrick, 2015). More significantly, Mattern, Radunzel, and Westrick assert that the ACT college readiness benchmarks were not developed to assess "a student's readiness for a specific college major or career field" (p. 5). Using hierarchical logistic regression on ACT science scores and grades from STEM-identified college science courses of nearly 70,000 students, Mattern, Radunzel, and Westrick found that an ACT science score of 25 (as opposed to the science college readiness benchmark of 23) resulted in students having an approximately 50% chance of earning a B or better in a first-year STEM science course in college. STEM science readiness benchmarks were not identified for the EXPLORE and PLAN tests, however.

For this study, higher percentages of all students and historically underserved students in the Physics First science sequence (30% and 14%, respectively) attained the STEM science college readiness benchmark of 25 developed by Mattern, Radunzel, and Westrick compared to their peers in a traditional science sequence (28% and 12%, respectively).

Finally, as with all studies conducted in a single district, the findings of this study “should be tempered with the understanding that successful change within educational settings is context-dependent” (Gaubatz , 2013, p. 25). I now attempt to put these findings in context as well as identify the implications of the results for policy and practice.

CHAPTER 5 DISCUSSION/CONCLUSION

The purpose of this study was to determine if the science college readiness and interest in STEM of historically underserved students can be improved by implementing a system-wide, inquiry-based high school science sequence with modeling comprised of coursework in physics, chemistry, and biology. This Physics First approach to high school science seeks to close gaps in content coverage experienced by historically underserved students by ensuring all students are exposed to the same content in physics, chemistry, and biology. Incorporating inquiry and modeling in this approach develops the academic skills of all students through active learning and the construction of new knowledge; thereby closing gaps in content exposure and emphasis more frequently experienced by historically underserved students compared to their White, economically advantaged peers whose first language is English.

A quantitative retrospective cohort observational study (Hoffmann & Lim, 2007; Mann, 2003) addressed the primary research question: do historically underserved students in a Physics First science sequence have 1) higher science college-readiness test scores, 2) higher rates of science college- and career-readiness, and 3) greater interest in STEM careers in grade 11 compared to their peers in a traditional science sequence? Multiple linear regression was used to compare 11th grade ACT science scores of students in a traditional science sequence and students in a Physics First sequence after controlling for 8th grade science scores and demographic differences between the two groups. Binary logistic regression was used to calculate odds ratios for 11th grade science college readiness status and interest in STEM of students in a Physics First sequence compared to

students in a traditional science sequence, again taking into account 8th grade science college readiness status or STEM interest as well as any demographic differences.

Varying degrees of positive effects for students in the Physics First with modeling sequence were observed on all three measures.

Synthesis of Findings

Students experiencing a Physics First with modeling science sequence had 11th grade science scores .74 points higher compared to their counterparts in a traditional science sequence. The differences in the mean scores represent between 18% and 42% of the mean gain science scores between grades 8 and 11 of students in traditional science instruction. The effect sizes of the increased science scores for students in the Physics First sequence was .06, representing a small treatment effect. Bloom et al. (2008) report the average annual gain in effect size on nationally-normed science tests is .19 between grades 9 and 10 and .15 between grades 10 and 11. Students experiencing a Physics First with modeling science sequence were 1.28 times as likely to meet the college- and career-readiness benchmark on the ACT science test as their peers experiencing a traditional science sequence. Students experiencing a Physics First with modeling science sequence were 1.37 times as likely to express interest in STEM in grade 11 as their peers experiencing a traditional science sequence. These three findings apply to students from historically underserved populations with one exception: the increase in 11th grade science scores for economically disadvantaged students enrolled in Physics First is .29 points higher than economically disadvantaged students in the traditional science sequence, compared to .74 points for other student populations. These three findings, summarized in Figure 8, suggest that a Physics First science sequence with

modeling better prepares historically underserved students for STEM careers than a traditional science sequence.

ACT Science Score	<ul style="list-style-type: none"> • .29 points higher for economically disadvantaged student in Physics First • .74 points higher for all other students in Physics First
Science College Readiness Status	<ul style="list-style-type: none"> • Compared to students in a traditional science sequence, Physics First students are 1.28 times as likely to meet the ACT science college- and career-readiness benchmark
STEM Interest	<ul style="list-style-type: none"> • Compared to students in a traditional science sequence, Physics First students are 1.37 times as likely to express interest in STEM

Figure 8. Grade 11 outcomes and effects of Physics First.

The Larger Context

For all students to meet the expectations of the Next Generation Science Standards (NGSS), educators must improve students' opportunity to learn by focusing on "what and how well students are taught in classrooms" (Herman, 2007, p. 4). Darling-Hammond (2010) asserts that "unequal access to high-level courses and challenging curriculum explains much of the difference in achievement between minority students and White students" (p.52). After analyzing the 2012 Programme for International Student Assessment (PISA) results, Schmidt et al. (2015) conclude that "any serious effort to reduce educational inequalities must address unequal content coverage within schools" (p. 381). The implementation of a common Physics First science sequence with modeling consisting of common units of instruction aligned to the Next Generation Science Standards is an effort to close the opportunity to learn gap experienced by

historically underserved students by reducing variation in content coverage both within and across schools in this district. In this district, students in a traditional science sequence could choose from a multitude of science course sequences. In the Physics First approach, all students were scheduled into a common sequence of physics, chemistry, and biology.

Reducing opportunity to learn gaps extends beyond content coverage (Conrad-Curry, 2011). Closing opportunity to learn gaps requires competent teachers who are informed by the research on best practices and provide excellent instruction (Conrad-Curry, 2011). This excellent instruction is captured by Darling-Hammond (2010):

Decades of research have shown that teachers who produce high levels of learning for initially lower- and higher-achieving students alike provide active learning opportunities involving student collaboration and many uses of oral and written language, connect to students' prior knowledge and experiences, provide hands-on learning opportunities, and engage students' higher order thought processes. (p.55)

Providing students with richer learning, inquiry-based instruction contributes to a socially-just pedagogy described by LeBlanc and Larke (2011), Moje (2007) and Thadani et al. (2010). Implementing a common science sequence for all students beginning with physics and grounded in inquiry and modeling (a Physics First approach) can close opportunity to learn (OTL) gaps experienced by traditionally underserved students in terms of both content and instructional quality while incorporating elements of a socially-just pedagogy. Lee (2005) asserts that when historically underserved students “are provided with equitable learning opportunities in school or in their communities, they demonstrate academic achievement, interest, and agency” (p. 438). The results of this study quantify improved science achievement and science college- and career-readiness

in grade 11 for historically underserved students experiencing a Physics First science sequence with modeling (after controlling for initial differences). Nonetheless, large gaps in 11th grade science achievement and college- and career-readiness remain between historically underserved students and their peers who are not historically underserved. These achievement gaps were evident in grade 8 as well. Disappointingly, no evidence from this study suggested the Physics First approach narrowed achievement gaps. Rather, all student groups benefitted equally (with the notable exception of less improvement in 11th grade science scores for economically disadvantaged students). In contrast, interest in STEM by historically underserved students is similar to the STEM interest of their peers who are not historically underserved. The implementation of a Physics First science sequence with modeling boosted STEM interest for all student groups, including historically underserved students, compared to their peers experiencing a traditional science sequence. Increasing interest in STEM and the science achievement of historically underserved students better positions these students to pursue further study and careers in STEM (Radunzel et al., 2016).

Implications

Given the small positive effects on 11th grade science scores, science college- and career-readiness, and interest in STEM for historically underserved students in the Physics First science sequence with modeling, I offer recommendations for policy and practice as well as for further study.

Recommendations for policy and practice. Academic preparation to succeed in post-secondary coursework is an essential component of college and career readiness (Harvey, Slate, Moore, Barnes, & Martinez-Garcia, 2013). Unfortunately, historically

underserved students continue to have less access to high quality academic preparation for post-secondary coursework leading to STEM degrees and careers (Berry et al., 2013; Tate, 2008). Implementing a common science sequence for all students beginning with physics and grounded in inquiry and modeling (a Physics First approach) can close opportunity to learn (OTL) gaps experienced by traditionally underserved high school students in terms of both content and instructional quality. However, as documented in this study and elsewhere, historically underserved students enter 9th grade with lower science achievement scores than their White and Asian peers (LeBlanc & Larke, 2011; Schmidt et al., 2015). In addition to closing opportunity to learn gaps at the high school level, better academic preparation of students in elementary and middle school is crucial for increasing the number of historically underserved students prepared for successful pursuit of STEM careers (Venkataraman, Riordan, & Olson, 2010). As noted by Bair and Bair (2014), students who arrive at high school lacking the prerequisite skills for successfully attaining the NGSS need additional supports and more time for learning. Findings from this study suggest this may be especially true for economically disadvantaged students. A first step for districts would be to identify the supports less prepared students need to be successful in a Physics First sequence. These supports may extend beyond science to mathematics. Districts implementing Physics First should then consider how, and in what forms, these supports will be made available to less prepared students.

Closing opportunity to learn gaps requires more than placing all students in a common high school sequence. As discussed in Chapter 2, quality instruction is critical

(Darling-Hammond, 2010; Haberman, 1991; Ladson-Billings, 1995). High quality professional development that enhances teachers' understanding of the nature of science and strengthens both the content knowledge and the pedagogical knowledge of how to effectively teach Physics First courses using inquiry and modeling methods is required (Asghar et al., 2012; Gibson & Brooks, 2012; Kesson & Henderson, 2010; McGee et al., 2013; National High School Center, 2008; Penuel et al., 2007). In addition, districts should assess and address the professional development needs of their science teachers related to effectively instructing students from historically underserved populations. Recall from Chapter 1 that Black and Hispanic/Latino students are underrepresented in physics classrooms (White & Tesfaye, 2011). Although collaborative scientific inquiry and modeling in this Physics First approach are culturally responsive instructional approaches for students from historically underserved populations (Kanter & Konstantopolous, 2010; Lee & Buxton, 2011; Patchen & Cox-Petersen, 2008); teachers must also believe all students from historically underserved populations are capable of successfully learning physics (and chemistry and biology) (Lakshmanan et al., 2011). Professional development to help science teachers leverage nontraditional funds of knowledge of their historically underserved students and build stronger understanding of and relationships with students of different cultures and backgrounds should be provided when needed (Banks et al., 2001; Lee & Buxton, 2011; Tan & Barton, 2010; Yerrick, Schiller, & Reisfeld, 2011). Multiple studies have found that students with better teacher-student relationships have higher student achievement (Gehlbach, Brinkworth, King, Hsu, McIntyre, & Rogers, 2016).

Recommendations for further study. This study has quantified the effects of implementing Physics First on the science college- and career-readiness and STEM interest of historically underserved students. Qualitative studies to better understand how and why a Physics First approach produced positive results for historically underserved students could be of value (Maxwell, 2004). While students in the Physics First with modeling sequence are more likely to be interested in STEM in grade 11, do students express more confidence in their abilities to learn science? And while historically underserved students are interested in STEM at similar levels as their non-historically underserved peers, do historically underserved students envision their success in a future STEM career? Further, how does the implementation of a Physics First sequence with modeling influence advanced science course taking by historically underserved students in grade 12?

Assessing teacher beliefs about the implementation of Physics First may also be informative. Do “teachers see a tension between providing a strong education for the able and willing students and at the same time providing for the uninterested or less able students” (Anderson in Larkin, et al., 2014, p. 828)? To what extent are “teacher beliefs about the limitations of their students in terms of ‘ability’ or ‘maturity’ an obstacle” to more student-centered approaches to instruction of scientific inquiry and modeling in the Physics First sequence (Wallace & Kang, 2004, p. 940)? To what extent are teacher beliefs and practices congruent with the changes in science instruction called for in the Next Generation Science Standards and embedded in this Physics First approach (Januszyk, Miller, & Lee, 2016)? And how do teachers respond pedagogically when

faced with perceived or real challenges of teaching all students physics, chemistry, and biology (Prime & Miranda, 2006)? Answering these questions might help explain the smaller positive effect of Physics First for economically disadvantaged students than for other student groups. Findings from such qualitative studies addressing these questions could also assist districts considering or implementing a Physics First approach.

Given that there were significant differences in 11th grade science test scores and interest in STEM between students in a Physics First science sequence with modeling and their peers who experienced traditional science instruction, additional quantitative analysis could yield additional insights. One natural area for further exploration would be to assess the impact of a Physics First approach on advanced science course taking of historically underserved students. Students who take advanced science courses, such as Advanced Placement or International Baccalaureate are more likely to enter college prepared for science coursework and to major in STEM (Klopfenstein, 2004; Radunzel, Mattern, & Westrick, 2016; Tai, et al., 2006; Trusty, 2002). A second avenue of inquiry would be to assess the relative contribution of each of the three courses of the Physics First sequence to gain in achievement and interest in STEM. Third, this study was conducted in a district with ten high schools. A number of the schools are large enough that school level effects could be explored and used by the district to identify schools where particularly effective instructional is occurring as well as schools where additional professional development or support for the change process would be beneficial. Coupling school level results with qualitative studies on factors affecting implementation

and success might assist other district in developing a plan to implement Physics First with modeling.

An examination of the relationship between science achievement and math achievement could also be of value. To what extent are mathematics skills as measured by the 8th grade EXPLORE test acting as a gatekeeper to student achievement in high school science? Is the lower effect on 11th grade science test scores for economically disadvantaged students in this study correlated with mathematics achievement prior to or during high school? Also, do students in a Physics First sequence have better math skills in 11th grade compared to their peers in a traditional science sequence controlling for any differences in 8th grade math test scores? The emphasis on the development and use of mathematical models (e.g., linear, quadratic, inverse) and emphasis on conceptual understanding in the Physics First approach can deepen students understanding of these mathematical relationships (Hill, 2013; O'Brien & Thompson, 2009). Glasser (2012) provides evidence of improved PSAT math scores for student in a Physics First sequence, however the study was conducted in a single private high school with a very small sample size.

Additional quantitative analysis of results by gender may also be informative. Women are also identified as a historically underserved population in STEM degree attainment and employment, particularly in physics, computer science, and engineering (Beede, Julian, Langdon, McKittrick, Khan, & Doms, 2011; Cheryan, Ziegler, Montoya, & Jiang, 2017). Did the increases in STEM interest and science college readiness of students in this Physics First implementation benefit male and female students equally?

Finally, this study examined the effects of implementing a Physics First on the science achievement and STEM interest of the first two cohorts of students in the sequence. As noted by Fullan (1994), “change involves learning to do something new” (p. 2843). Asking teachers to implementing a new curriculum and associated methods of teaching, in this case inquiry and modeling, may engender feelings in some teachers of being de-skilled (Altrichter, 2005) despite professional development provided by the district. Altrichter (2005) also recognized that teacher learning in the context of curriculum implementation extends over time. Analysis of student achievement and STEM interest from the third and even fourth cohorts of students in Physics First would assess the effects of long term implementation and could also be used by the district to both monitor and improve the adopted Physics First approach.

Improving the STEM readiness of students from historically underserved groups is a moral and economic imperative (Levin, 2009; National Academies, 2007). The purpose of this study was to determine if the science college readiness of historically underserved students could be improved by implementing this Physics First approach for all students. A retrospective cohort observational study using multiple linear regression and binary logistic regression assessed the differences in 11th grade science college- and career-readiness test scores, rates of science college- and career-readiness, and interest in STEM between historically underserved students in a Physics First science sequence and their peers in a traditional sequence. The results of this study found implementing a high school science sequence beginning with physics and centered on developing conceptual understanding through inquiry labs and modeling had small positive effects on science

college- and career-readiness and interest in STEM careers for historically underserved students. While findings from a study conducted in a single district “should be tempered with the understanding that successful change within educational settings is context-dependent” (Gaubatz , 2013, p. 25), this study adds to the limited literature on the effectiveness of a Physics First approach (Glasser, 2012) and confirms positive effects found in the study by Dye et al.(2013). This study also breaks new ground by quantifying outcomes of Physics First for historically underserved students, a topic which has been unexplored to date, but is more important than ever in an era of increasing racial and ethnic diversity, income inequality, technological advancement and global competition.

REFERENCES

- Abedi, J., & Herman, J. (2010). Assessing English language learners' opportunity to learn mathematics: Issues and limitations. *Teachers College Record*, 112(3), 723–746.
- ACT. (2009a). *The ACT interest inventory and the world-of-work map*. Iowa City, IA: author. Retrieved March 12, 2016 from http://www.act.org/content/dam/act/unsecured/documents/interest_inventory.pdf
- ACT. (2009b). *How much growth toward college readiness is reasonable to expect in high school?* Iowa City, IA: author. Retrieved November 27, 2016 from <http://www.act.org/content/dam/act/unsecured/documents/ReasonableGrowth.pdf>
- ACT. (2013a). *ACT EXPLORE technical manual*. Iowa City, IA: author. Retrieved March 12, 2016 from <https://www.act.org/content/dam/act/unsecured/documents/Explore-TechManual.pdf>
- ACT. (2013b). *ACT PLAN technical manual*. Iowa City, IA: author. Retrieved March 12, 2016 from <http://forms.act.org/plan/pdf/PlanTechnicalManual.pdf>
- ACT. (2013c). *Information brief 2013-22: Science courses taken and ACT college readiness benchmark performance in science*. Retrieved from <http://www.act.org/research/researchers/briefs/2013-22.html>
- ACT. (2013d). *What are the ACT college readiness benchmarks? Information Brief*. Retrieved from ERIC. <https://eric.ed.gov/?id=ED546859>
- ACT. (2014). *ACT technical manual*. Iowa City, IA: author. Retrieved March 12, 2016 from http://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf
- ACT. (2015). *The condition of STEM 2015*. Retrieved March 12, 2016 from <http://www.act.org/stemcondition/13/pdf/National-STEM-Report-2013.pdf>
- Allen, J. (2013). *Updating the ACT college readiness benchmarks*. (ACT Research Report No. 2013-6). Iowa City, IA: ACT, Inc.
- Allen, J., & Sconing, J. (2005). *Using ACT assessment scores to set benchmarks for college readiness*. (ACT Research Report No. 2005-3). Iowa City, IA: ACT, Inc.

- Altrichter, H. (2005). Curriculum implementation—limiting and facilitating factors. In P. Nentwig & D. Waddington (Eds.), *Context based learning of science*, (pp. 35-62). Waxmann: Münster.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anderson, C. J. & Rutkowski, L. (2008). Multinomial logistic regression. In Osborne, J. W. (Ed.), *Best practices in quantitative methods* (pp. 390 – 409). Thousand Oaks, CA.: Sage Publications.
- Anderson, D. M. (2008). Continuing professional development: Chemistry for non-specialists - A report. *Education in Science*, (226), 16.
- Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management*, 15(4), 623–645.
- Asghar, A., Ellington, R., Rice, E., Johnson, F., & Prime, G. M. (2012). Supporting STEM education in secondary science contexts. *Interdisciplinary Journal of Problem-Based Learning*, 6(2), 85–125.
- Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38(9), 665–676.
- Bailey, M. J., & Dynarski, S. M. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (p. 117-132). New York, NY: Russell Sage Foundation & Spencer Foundation.
- Bair, M. A., & Bair, D. E. (2014). Failure, the next generation: Why rigorous standards are not sufficient to improve science learning. *International Journal of Education Policy & Leadership*, 9(5), 1–12.
- Banks, J. A., Cookson, P., Gay, G., Hawley, W. D., Irvine, J. J., Nieto, S., Schofield, J. W., Stephan, W. G. (2001). Diversity within unity: Essential principles for teaching and learning in a multicultural society. *Phi Delta Kappan*, 83(3), 196–203.

- Bardeen, M. G., & Lederman, L. M. (1998). Coherence in science education. *Science*, 281(5374), 178-179.
- Baum, S., Ma, J., & Payea, K. (2010). Education pays: The benefits of higher education for individuals and society. Washington, D.C.: *The College Board*.
- Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11).
- Bermudez, J. V. (2014). *Examining the effects of physics second on high school science achievement* (Master's thesis, California State University, Long Beach). Retrieved from ProQuest Dissertations & Theses Global. (1530197157)
- Berry, R. Q., Ellis, M., & Hughes, S. (2014). Examining a history of failed reforms and recent stories of success: Mathematics education and Black learners of mathematics in the United States. *Race Ethnicity and Education*, 17(4), 540–568.
- Bettinger, E. P., Evans, B. J., & Pope, D. G. (2011). *Improving college performance and retention the easy way: Unpacking the ACT exam*. NBER Working Paper No. 17119. National Bureau of Economic Research. Retrieved from ERIC. <http://eric.ed.gov/?id=ED521075>
- Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4), 289–328.
- Boone, S. (2006). Textbooks: Influences on selection. *Digest of Gifted Research* (8/28/2006) Retrieved from <http://tip.duke.edu/node/779>.
- Boscardin, C. K., Aguirre-Muñoz, Z., Stoker, G., Kim, J., Kim, M., & Lee, J. (2005). Relationship between opportunity to learn and student performance on English and algebra assessments. *Educational Assessment*, 10(4), 307–332.
- Bouma, C. E. (2013). *Physics First: Impact on SAT math scores*. (Doctoral dissertation, Loyola Marymount University). Retrieved from ProQuest Dissertations & Theses Global. (1499247187)
- Briggs, D. C. (2008). Synthesizing causal inferences. *Educational Researcher*, 37(1), 15–22.
- Brown, R., Brown, J., Reardon, K., & Merrill, C. (2011). Understanding STEM: Current perceptions. *Technology and Engineering Teacher*, 70(6), 5-9.

- Buddin, R. & Croft, M. (2014) Do stricter high school graduation requirements improve college readiness? ACT Working Paper WP-2014-1. Retrieved from <http://www.act.org/research/papers/pdf/wp-2014-1.pdf>
- Burris, C. C., & Welner, K. G. (2005). Closing the achievement gap by detracking. *Phi Delta Kappan*, 86(8), 594–598.
- Burnett, F. (2010). Improving students college and career readiness. *Counseling Today*, 53(4), 42-43.
- Cakir, M. (2008). Constructivist approaches to learning in science and their implications for science pedagogy: A literature review. *International Journal of Environmental & Science Education*, 3(4), 193–206.
- Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal*, 42(2), 305–328.
- Camburn, E. M. (2010). Embedded teacher learning opportunities as a site for reflective practice: An exploratory study. *American Journal of Education*, 116(4), 463–489.
- Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio Books.
- The Center for Education Policy Analysis. (2008). *Learning about science, technology, engineering, and mathematics*. Retrieved April 20, 2014 from <http://www.ucdenver.edu/academics/colleges/SPA/BuechnerInstitute/Research/Documents/CEPASTEMReportFinal.pdf>
- Center for Public Education. (2014). Understanding the common core standards. *Education Digest*, 79(8), 16–21.
- Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143(1), 1–35.
- Children's Defense Fund (2012). *The state of America's children handbook*. Retrieved from <http://www.childrensdefense.org/child-research-data-publications/data/soac-2012-handbook.pdf>
- Cochran, W. G. (1983). *Planning and analysis of observational studies*. New York, NY: Wiley.
- Conley, D. T. (2010). *College and career ready: Helping all students succeed beyond high school*. San Francisco, CA: Jossey-Bass.

- Conrad-Curry, D. (2011). A four-year study of ACT reading results: Achievement trends among eleventh-grade boys and girls in a midwestern state. *Journal of Education*, 191(3), 27–37.
- Contreras, F. (2011). Strengthening the bridge to higher education for academically promising underrepresented students. *Journal of Advanced Academics*, 22, 500–526.
- Cowan Pitre, C. (2014). Improving African American student outcomes: Understanding educational achievement and strategies to close opportunity gaps. *Western Journal of Black Studies*, 38(4), 209–217.
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.
- Day, S. X., Rounds, J., & Swaney, K. (1998). The structure of vocational interests for diverse racial-ethnic groups. *Psychological Science*, 9(1), 40–44.
- Deming, D., & Dynarski, S. (2009). *Into college, out of poverty? Policies to increase the postsecondary attainment of the poor* (No. w15387). National Bureau of Economic Research.
- Duncan, G. J., & Murnane, R. J. (2014). Growing income inequality threatens American education. *Phi Delta Kappan*, 95(6), 8–14.
- Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34(9), 917–928.
- Dye, J., Cheatham, T., Rowell, G. H., Barlow, A. T., & Carlton, R. (2013). The impact of modeling instruction within the inverted curriculum on student achievement in science. *Electronic Journal of Science Education*, 17(2), 1–19.
- Ewald, G., Hickman, J. B., Hickman, P., & Myers, F. (2005). Physics First: The right-side-up science sequence. *The Physics Teacher*, 43(5), 319–320.
- Field, A. P. (2016). *Discovering statistics using IBM SPSS statistics*. Los Angeles : Sage.
- Fields, E. T., Levy, A. J., Karelitz, T. M., Martinez-Gudapakkam, A., & Jablonski, E. (2012). The science of professional development. *Phi Delta Kappan*, 93(8), 44.
- Finn, C. E., & Ravitch, D. (2004). *The mad, mad world of textbook adoption*. Washington, DC: Thomas B. Fordham Institute.

- Freeman, C.W. (2008). China's real three challenges to the U.S. April 20, 2014 from <http://www.theglobalist.com/StoryId.aspx?StoryId=5770>
- Fullan, M. (1994). Implementation of innovations. In: T. Husén, & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education*. (pp. 2839-2847). Oxford, England: Pergamon.
- Fullan, M. (2002). The change leader. *Educational Leadership*, 59(8), 16.
- Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation & Policy Analysis*, 19(4), 325–338.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Kwang Suk Yoon. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945.
- Gasbarra, P. & Johnson, J. (2008). *Out before the game begins: Hispanic leaders talk about what's needed to bring more Hispanic youngsters into science, technology, and math professions*. Retrieved April 20, 2014 from <http://www.publicagenda.org/files/pdf/outbefore.PDF>
- Gaubatz, J. (2013). Evaluation of a secondary school science program inversion: Moving from a traditional to a modified-PCB sequence. *Science Educator*, 22(1), 19–25.
- Gehlbach, H., Brinkworth, M. E., King, A. M., Hsu, L. M., McIntyre, J., & Rogers, T. (2016). Creating birds of similar feathers: Leveraging similarity to improve teacher-student relationships and academic achievement, *Journal of Educational Psychology*, 108(3), 342–352.
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45(8), 922–939.
- Gerlach, J. (2012). STEM: defying a simple definition. *NSTA Reports* (April 11, 2012), p. 3. Arlington, VA: National Science Teachers Association.
- Gibson, S. E., & Brooks, C. (2012). Teachers' perspectives on the effectiveness of a locally planned professional development program for implementing new curriculum. *Teacher Development*, 16(1), 1–23.

- Gilbert, A., & Yerrick, R. (2001). Same school, separate worlds: a sociocultural study of identity, resistance, and negotiation in a rural, lower track science classroom. *Journal of Research in Science Teaching*, 38(5), 574–598.
- Glasser, H. M. (2012). The numbers speak: Physics First supports math performance. *The Physics Teacher*, 50(1), 53–55.
- Goodman, R. & Etkina, E. (2008). Squaring the circle: A mathematically rigorous physics first. *The Physics Teacher*, 46(4), 222–227.
- Haber-Schaim, U. (1984). In my opinion: High school physics should be taught before chemistry and biology. *The Physics Teacher*, 22(5), 330–332.
- Haberman, M. (1991). The pedagogy of poverty versus good teaching. *The Phi Delta Kappan*, 73(4), 290–294.
- Harvey, D. W., Slate, J. R., Moore, G. W., Barnes, W., & Martinez-Garcia, C. (2013). College readiness gaps: A review of the literature. *Journal of Education Research*, 7(3), 181–204.
- Heifetz, R. A., & Linsky, M. (2004). When leadership spells danger. *Educational Leadership*, 61(7), 33–37.
- Herman, J. L. (2007). Accountability and assessment: Is public interest in K-12 education being served? CRESST Report 728. National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from ERIC. <http://eric.ed.gov/?id=ED499421>
- Herman, J. L., Klein, D. C. D., & Abedi, J. (2000). Assessing students' opportunity to learn: Teacher and student perspectives. *Educational Measurement: Issues and Practice*, 19(4), 16–24.
- Hernandez, P., Schultz, P., Estrada, M., Woodcock, A., & Chance, R. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *Journal of Educational Psychology* 105 (1), 89–107.
- High School Committee of the American Association of Physics Teachers (2009, January). Physics First: An Informational Guide for Teachers, School Administrators, Parents, Scientists, and the Public, American Association of Physics Teachers. Retrieved November 23, 2014 from http://www.aapt.org/resources/upload/phys_first.pdf
- Hill, B. (2013). The patterns approach. *Science Teacher*, 80(3), 38–42.

- Hoffmann, R. G., & Lim, H. J. (2007). Observational study design. In W. T. Ambrosius (Ed.), *Topics in Biostatistics* (pp. 19–31). Totowa, NJ: Humana Press. Retrieved from http://dx.doi.org/10.1007/978-1-59745-530-5_2
- Humes, K.R., Jones, N. A., & Ramirez, R. R. (2011, March). *Overview of race and Hispanic origin: 2010*. Washington, D.C.: U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. Retrieved May 19, 2014 from <http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>
- Hunzicker, J. (2012). Professional development and job-embedded collaboration: How teachers learn to exercise leadership. *Professional Development in Education*, 38(2), 267–289.
- Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation : A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences*. San Diego, CA: EdITS Publishers.
- Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling instruction: An effective model for science education. *Science Educator*, 17(1), 10–17.
- Januszyk, R., Miller, E. C., & Lee, O. (2016). Addressing student diversity and equity. *Science & Children*, 53(8), 28–31.
- Jeynes, W. H. (2015). A meta-analysis on the factors that best reduce the achievement gap. *Education & Urban Society*, 47(5), 523–554.
- Ji, P., DuBois, D. L., Flay, B. R., & Brechling, V. (2008). “Congratulations, you have been randomized into the control group! (?)”: Issues to consider when recruiting schools for matched-pair randomized control trials of prevention programs. *Journal of School Health*, 78(3), 131–139.
- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, 30(2), 3–13.
- Johnson, C. C. (2009). An examination of effective practices: Moving toward elimination of achievement gaps in science. *Journal of Science Teacher Education*, 20(3), 287–306.
- Kannapel, P. J., & Clements, S. K. (2005, February). *Inside the black box of high-performing high-poverty schools*. Lexington, KY: The Prichard Committee for Academic Excellence. Retrieved April 27, 2014 from <http://www.prichardcommittee.org/wp-content/uploads/2013/02/Inside-the-Black-Box.pdf>

- Kanter, D. E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content knowledge and inquiry-based practices. *Science Education*, 94(5), 855-887.
- Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17(2), 137-152.
- Kesson, K. R., & Henderson, J. G. (2010). Reconceptualizing professional development for curriculum leadership: Inspired by John Dewey and informed by Alain Badiou. *Educational Philosophy and Theory*, 42(2), 213-229.
- Klopfenstein, K. (2004). Advanced Placement: do minorities have equal opportunity? *Economics of Education Review*, 23(2), 115-131.
- Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *Science & Children*, 49(7), 10-13.
- Krathwohl, D.R. (2009). *Methods of educational and social science research: The logic of methods*. Long Grove, IL: Waveland Press, Inc.
- Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory into Practice*, 34(3), 159-165.
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, 35(7), 3-12.
- Lakshmanan, A., Heath, B. P., Perlmutter, A., & Elder, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *Journal of Research in Science Teaching*, 48(5), 534-551.
- Larkin, D. B., Seyforth, S. C., & Lasky, H. J. (2009). Implementing and sustaining science curriculum reform: A study of leadership practices among teachers within a high school science department. *Journal of Research in Science Teaching*, 46(7), 813-835.
- LeBlanc, J., & Larke, P. J. (2011). Culturally responsive teaching in science. *National Forum of Multicultural Issues Journal*, 8(1), 40-51.

- Lederman, L. M. (1998). *ARISE: American renaissance in science education*. Fermilab-TM-2051. Batavia, IL: Fermi National Accelerator Lab. Retrieved from: <http://lss.fnal.gov/archive/1998/tm/TM-2051.pdf>
- Lederman, N.G., Lederman, J.S., & Antink, A. (2013). Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. *International Journal of Education in Mathematics, Science and Technology*, 1(3), 138-147.
- Lee, O. (2005). Science education and student diversity: Synthesis and research agenda. *Journal Of Education For Students Placed At Risk*, 10(4), 433-443.
- Lee, O., & Buxton, C. (2011). Engaging culturally and linguistically diverse students in learning science. *Theory Into Practice*, 50(4), 277–284.
- Levin, H. M. (2009). The economic payoff to investing in educational justice. *Educational Researcher*, 38(1), 5-20.
- Liang, L., Fulmer, G., Majerich, D., Clevenstine, R., & Howanski, R. (2012). The effects of a model-based physics curriculum program with a physics first approach: a causal-comparative study. *Journal of Science Education & Technology*, 21(1), 114–124.
- Livanis, O. (2006). What was I thinking? *The Physics Teacher*, 44(5), 280–282.
- Loveless, T. (1999). Will tracking reform promote social equity? *Educational Leadership*, 56(7), 28-32.
- Luft, J. A., da Cunha, T., & Allison, A. (1998, April). *Increasing the participation of minority students in science: A study of two teachers*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, San Diego, CA. Retrieved from ERIC. <http://eric.ed.gov/?id=ED418944>
- Lutz, S. T., & Huitt, W. G. (2004). Connecting cognitive development and constructivism: Implications from theory for instruction and assessment. *Constructivism in the Human Sciences*, 9(1), 67-90.
- Mann, C. J. (2003). Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emergency Medicine Journal*, 20(1), 54-60.
- Marshall, J. C., & Alston, D. M. (2014). Effective, sustained inquiry-based instruction promotes higher science proficiency among all groups: A 5-year analysis. *Journal of Science Teacher Education*, 25(7), 807–821.

- Maruyama, G. (2012). Assessing college readiness: Should we be satisfied with ACT or other threshold scores? *Educational Researcher*, 41(7), 252–261.
- Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tali, R. T. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching*, 41(10), 1063–1080.
- Mary, M. T. (2015). *An investigation of the impact of science course sequencing on student performance in high school science and math* (Doctoral dissertation, The University of Texas at San Antonio). Retrieved from ProQuest Dissertations & Theses Global. (1719279467)
- Mattern, K., Radunzel, J., & Westrick, P. (2015). *Development of STEM readiness benchmarks to assist career and educational decision making*. ACT Research Report 2015-3. Iowa City, IA: ACT, Inc.
- Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33(2), 3–11.
- Maxwell, J.A. (2013). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage Publications, Inc.
- McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, 17(3), 305–322.
- McGee, J. R., Wang, C., & Drew, P. (2013). Guiding teachers in the use of a standards-based mathematics curriculum: Teacher perceptions and subsequent instructional practices after an intensive professional development program. *School Science and Mathematics*, 113(1), 16–28.
- McNamee R. (2005). Regression modelling and other methods to control confounding. *Occupational and Environmental Medicine* 62(7), 500–506.
- McNeish, D., Radunzel, J., & Sanchez, E.I. (2015). A multidimensional perspective of college readiness: Relating student and school characteristics to performance on the ACT. *ACT Research Report Series*, RR2015-6. Iowa City, IA: ACT, Inc.
- Mervis, J. (1998). U.S. tries variations on high school curriculum. *Science*, 281, 161,163.
- Milloy, M. (2003). Want to be a scientist? *NEA Today*, 21(8), 14.
- Milner, H. R. (2012). Beyond a test score: Explaining opportunity gaps in educational practice. *Journal of Black Studies*, 43(6), 693–718.

- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction--what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496.
- Moje, E. B. (2007). Developing socially just subject-matter instruction: A review of the literature on disciplinary literacy teaching. *Review of Research in Education*, 31(1), 1–44.
- Montgomery, N., Allensworth, E. M., & Correa, M. (2010). *Passing through science: The effects of raising graduation requirements in science on course-taking and academic achievement in Chicago. Research Report* (No. 978-0-9814-6047-5). Consortium on Chicago School Research. Chicago: IL. Retrieved from ERIC: <http://eric.ed.gov/?id=ED509706>
- National Academies of Science. (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5*. Report from the Committee on Prospering in the Global Economy of the 21st Century. Washington; DC: National Academies Press.
- National Academies of Science. (2007). *Rising above the gathering storm*. Report from the Committee on Prospering in the Global Economy of the 21st Century. Washington; DC: National Academies Press.
- National Academies of Sciences, Engineering, and Medicine. (2015). *Science teachers learning: Enhancing opportunities, creating supportive contexts*. Committee on Strengthening Science Education through a Teacher Learning Continuum. Board on Science Education and Teacher Advisory Council, Division of Behavioral and Social Science and Education. Washington, DC: The National Academies Press.
- National High School Center. (2011). *Eight elements of high school improvement: A mapping framework*. National High School Center. Retrieved from ERIC: <https://eric.ed.gov/?id=ED521559>
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- National Science Teachers Association (NSTA). (2011). Six more states join effort to write next generation science standards. *NSTA Express*, (12/5/2011), 3.

- Noble, J., Davenport, M., Schiel, J., & Pommerich, M. (1999). *High school academic and noncognitive variables related to the ACT scores of racial/ethnic and gender groups* (ACT Research Report No. 99-6). Iowa City, IA: ACT.
- Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: Rand Corp. Retrieved from ERIC. <https://eric.ed.gov/?id=ED329615>
- Oakes, J., & Wells, A. S. (1998). Detracking for high student achievement. *Educational Leadership*, 55(6), 38-41.
- O'Brien, M. J., & Thompson, J. R. (2009). Effectiveness of ninth-grade physics in Maine: Conceptual understanding. *Physics Teacher*, 47(4), 234–239.
- Ornstein, A. C. (2010). Achievement gaps in education. *Society*, 47, 424–429.
- Osborne, J. W. (2008). *Best practices in quantitative methods*. Thousand Oaks, Calif.: Thousand Oaks, Calif. : Sage Publications.
- Ottmar, E. R., Konold, T. R., Berry, R. Q., Grissmer, D. W., & Cameron, C. E. (2013). Increasing equity and achievement in fifth grade mathematics: The contribution of content exposure. *School Science and Mathematics*, 113(7), 345–355.
- Pasero, S. (2001). The State of Physics-First Programs. *Project ARISE*.
- Patchen, T., & Cox-Petersen, A. (2008). Constructing cultural relevance in science: A case study of two elementary teachers. *Science Education*, 92(6), 994–1014.
- Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921–958.
- Peterson-Beeton, R. (2007). Minority students' perspectives on chemistry in an alternative high school. *The Qualitative Report*, 12(4), 705–728.
- Plano Clark, V. L. & Creswell, J. W. (2015). *Understanding research : A consumer's guide*. Upper Saddle River, N.J.: Merrill/Pearson Educational.
- Plunk, A. D., Tate, W. F., Bierut, L. J., & Grucza, R. A. (2014). Intended and unintended effects of state-mandated high school science and mathematics course graduation requirements on educational attainment. *Educational Researcher*, 20(10), 1–12.

- Popkin, G. (2009). "Physics First" battles for acceptance. *APS News*, 18(7) 3. Retrieved January 4, 2016 from <https://www.aps.org/publications/apsnews/200907/upload/July-2009.pdf>
- Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology From Bed to Bench*, 5(2), 79–83.
- Prediger, D. J., & Swaney, K. B. (2004). Work task dimensions underlying the world of work: Research results for diverse occupational databases. *Journal of Career Assessment*, 12(4), 440–459.
- Prime, G. M., & Miranda, R. J. (2006). Urban public high school teachers' beliefs about science learner characteristics. *Urban Education*, 41(5), 506–532.
- Putnam, R. D. (2015). *Our kids: The American dream in crisis*. New York : Simon & Schuster.
- Radcliffe, R. A. & Bos, B. (2013). Strategies to prepare middle school and high school students for college and career readiness. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 86(4), 136-141.
- Radunzel, J., Mattern, K., & Westrick, P. (2016). *The role of academic preparation and interest on STEM success*. *ACT Research Report Series, 2016 (8)*. ACT, Inc. Retrieved from <http://www.act.org/content/dam/act/unsecured/documents/5940-Research-Report-2016-8-Role-of-Academic-Preparation-and-Interest-on-STEM-Success.pdf>
- Radunzel, J., & Noble, J. (2013). *Differential effects on student demographic groups of using ACT college readiness assessment composite score, ACT benchmarks, and high school grade point average for predicting long-term college success through degree completion*. *ACT Research Report Series, 2013 (5)*. ACT, Inc. Retrieved from ERIC. <https://eric.ed.gov/?id=ED555596>
- Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership* 70(8), 10-16.
- Robelen, E. W. (2010). Panel moves toward "next generation" science standards. *Education Digest*, 76(2), 32–33.
- Robelen, E. W. (2013). Standards in science unveiled. *Education Week*, 32(28), 1, 12-13.
- Rotherham, A. J., & Willingham, D. (2009). 21st century skills: The challenges ahead. *Educational Leadership*, 67(1), 16–21.

- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020–1049.
- Sadler, P. M., & Tai, R. H. (2007). The two high-school pillars supporting college science. *Science*, 317(5837), 457–458.
- Sawyer, R. (2010). *Usefulness of high school average and ACT scores in making college admission decisions. ACT Research Report Series 2010-2*. ACT, Inc. Retrieved from ERIC. <https://eric.ed.gov/?id=ED527216>
- Sawyer, R. (2013). *Interpreting changes over time in high school average ACT college readiness assessment composite scores and ACT college readiness benchmark attainment rates. ACT Research Report Series, 2013 (9)*. ACT, Inc. Retrieved from ERIC. <https://eric.ed.gov/?id=ED555601>
- Schiller, K. S., Schmidt, W. H., Muller, C., & Houang, R. T. (2010). Hidden disparities: How courses and curricula shape opportunities in mathematics during high school. *Equity & Excellence in Education*, 43(4), 414–433.
- Schmidt, W. H., & Burroughs, N. A. (2013). Springing to life: How greater educational equality could grow from the common core mathematics standards. *American Educator*, 37(1), 2–9.
- Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality. *Educational Researcher*, 44(7), 371–386.
- Schmidt, W. H., & McKnight, C. C. (2012). *Inequality for all: The challenge of unequal opportunity in American schools*. New York, NY: Teachers College Press.
- Schmidt, W. H., Wang, H. C., & McKnight, C.C. (2005). Curriculum coherence: An examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, 37(5), 525–559.
- Schroeder, C. M., Scott, T. E., Tolson, H., Huang, T., & Lee, Y. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44(10), 1436–1460.
- Schultheis, E. H., & Kjervik, M. K. (2015). Data nuggets: Bringing real data into the classroom to unearth students' quantitative & inquiry skills. *The American Biology Teacher*, 77(1), 19–26.

- Schwartz, M. S., Sadler, P. M., Sonnert, G., & Tai, R. H. (2009). Depth versus breadth: How content coverage in high school science courses relates to later success in college science coursework. *Science Education*, 93(5), 798–826.
- Shanahan, M., Pedretti, E., DeCoito, I., & Baker, L. (2013). Exploring the responses of underrepresented students in science to an elementary classroom outreach program. *School Science & Mathematics*, 111(4), 131-142.
- Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14.
- Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation & Policy Analysis*, 31(4), 500–506.
- Soper, D.S. (2016). Post-hoc Statistical Power Calculator for Multiple Regression [Software]. Available from <http://www.danielsoper.com/statcalc>
- Sparks, S. D. (2013). A fifth of children remain poor, says census. *Education Week* 33(5), 5.
- Spence, D. S. (2007). A road map to college and career readiness. *Education Week*, 26(17), 59-65.
- State Educational Technology Directors Association. (2008). *Science, technology, engineering, and math (STEM)*. State Educational Technology Directors Association (SETDA). Retrieved from <http://www.setda.org/wp-content/uploads/2013/11/Science-Technology-Engineering-and-Mathematics-STEM-Report.pdf>
- Stevens, F. I. (1993). Applying an opportunity-to-learn conceptual framework to the investigation of the effects of teaching practices via secondary analyses of multiple- case-study summary data. *The Journal of Negro Education*, 62(3), 232-248. <http://doi.org/10.2307/2295463>
- Tai, R. H., Sadler, P. M., & Mintzes, J. J. (2006). Factors influencing college science success. *Journal of College Science Teaching*, 36(1), 52-56.
- Tan, E., & Barton, A. C. (2010). Transforming science learning and student participation in sixth grade science: A case study of a low-income, urban, racial minority classroom. *Equity & Excellence in Education*, 43(1), 38–55.

- Tate, W. (2008). The political economy of teacher quality in school mathematics: African American males, opportunity structures, politics, and method. *American Behavioral Scientist*, 51(7), 953–973.
- Taylor, J. A., Powell, J. C., Van Dusen, D. R., Schindler, B. J., Pearson, B., Lavine, D., & Bess, K. (2005). Curriculum reform and professional development in San Diego City Schools. *Physics Teacher*, 43(2), 102–106.
- Teitelbaum, P. (2003). The influence of high school graduation requirement policies in mathematics and science on student course-taking patterns and achievement. *Educational Evaluation & Policy Analysis*, 25(1), 31–57.
- Temple, J. A., & Reynolds, A. J. (2000). School mobility and achievement: Longitudinal findings from an urban cohort. *Journal of School Psychology*, 37(4), 355–377.
- Thadani, V., Cook, M., Griffis, K., Wise, J. A., & Blakely, A. (2010). The possibilities and limitations of curriculum-based science inquiry interventions for challenging the ‘pedagogy of poverty’. *Equity & Excellence in Education*, 43(1), 21–37.
- Trusty, J. (2002). Effects of high school course-taking and other variables on choice of science and mathematics college majors. *Journal of Counseling & Development*, 80(4), 464–474.
- Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, 6(4), 371–386.
- Tuckman, B. W. (1994). *Conducting educational research*. Orlando, FL: Harcourt Brace & Company.
- Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk (JESPAR)*, 12(3), 243–270.
- U.S. Census Bureau, Projected Population of the United States, by Race and Hispanic Origin: 2012 to 2060, December 2012.
<http://www.census.gov/newsroom/releases/archives/population/cb12-243.html>
- Venkataraman, B., Riordan, D. G., & Olson, S. (2010). Prepare and inspire: K-12 education in science, technology, engineering, and math (STEM) for America's future. *Washington, DC: President's Council of Advisors on Science and Technology*. Retrieved from:

<https://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-stemed-report.pdf>

- Voight, A., Shinn, M., & Nation, M. (2012). The longitudinal effects of residential mobility on the academic achievement of urban elementary and middle school students. *Educational Researcher*, 41(9), 385-392.
- Wallace, C. S., & Kang, N. H. (2004). An investigation of experienced secondary science teachers' beliefs about inquiry: An examination of competing belief sets. *Journal of Research in Science Teaching*, 41(9), 936-960.
- Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis*, 20(3), 137-156.
- Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment*, 20(1), 23-45.
- White, S., & Tesfaye, C., (2011). Under-represented minorities in high school physics: Results from the 2008-09 nationwide survey of high school physics teachers. Focus on. *Statistical Research Center of the American Institute of Physics*.
- Williams, C. A. (2009). *Impact of Physics First on PSAT achievement in science and math* (Doctoral dissertation, Aurora University). Retrieved from ProQuest Dissertations & Theses Global. (305168715)
- Yerrick, R., Schiller, J., & Reisfeld, J. (2011). "Who are you callin' expert?": Using student narratives to redefine expertise and advocacy lower track science. *Journal of Research in Science Teaching*, 48(1), 13-36.
- Zinth, K., & Dounay, J. (2006). *Mathematics and science education in the states. StateNotes*. Denver, CO: Education Commission of the States. Retrieved from ERIC. <https://eric.ed.gov/?id=ED493716>
- Zuniga, K., Olson, J. K., & Winter, M. (2005). Science education for rural Latino/a students: Course placement and success in science. *Journal of Research in Science Teaching*, 42(4), 376-402.

APPENDIX A CONTENT SPECIFICATIONS FOR THE ACT SCIENCE TEST

The ACT Science Test is a 40-item test drawing on content typically covered in high school science courses (ACT, 2014). The test emphasizes scientific reasoning skills over recall of scientific content. Materials are drawn from biology Earth/space science, physics, and chemistry with at least one passage, and no more than two passages, from each content area (ACT, 2014). Advanced knowledge in these four subjects is not required, but background knowledge acquired in general introductory science courses is necessary for some questions (ACT, 2014). Questions are presented in three formats: Data Representation, Research Summaries, and Conflicting Viewpoints (ACT, 2014). Data Representation (30% of questions) present students with graphics and tables similar to that found in science journals and texts to measure student skills in reading graphs, interpreting scatterplots, and interpreting information presented in tables, diagrams, and figures (ACT, 2014). Research Summaries (50% of questions) provides students with descriptions of one or more related experiments to assess student skills in design of experiments and the interpretation of experimental results (ACT, 2014). Conflicting Viewpoints (20% of questions) presents students several hypotheses or views based on differing premises or on incomplete data which are inconsistent with one another in order to assess student understanding, analysis, and comparison of alternative viewpoints or hypotheses (ACT, 2014).

APPENDIX B LINEAR REGRESSION MODEL

Model Summary^h

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics		
					R Square Change	F Change	Sig. F Change
1	.391 ^a	.153	.152	5.360	.153	230.403	.000
2	.451 ^b	.203	.202	5.198	.051	487.109	.000
3	.473 ^c	.224	.223	5.130	.021	206.960	.000
4	.747 ^d	.558	.558	3.870	.334	5811.246	.000
5	.749 ^e	.561	.561	3.858	.003	51.783	.000
6	.755 ^f	.570	.569	3.823	.008	18.411	.000
7	.755 ^g	.570	.569	3.822	.000	5.824	.016

a. Predictors: (Constant), MU, AI, PI, BL, AS, Latino

b. Predictors: (Constant), MU, AI, PI, BL, AS, Latino, EconDsvntgFg

c. Predictors: (Constant), MU, AI, PI, BL, AS, Latino, EconDsvntgFg, LEPPFg

d. Predictors: (Constant), MU, AI, PI, BL, AS, Latino, EconDsvntgFg, LEPPFg, Science Score

e. Predictors: (Constant), MU, AI, PI, BL, AS, Latino, EconDsvntgFg, LEPPFg, Science Score, Treat

f. Predictors: (Constant), MU, AI, PI, BL, AS, Latino, EconDsvntgFg, LEPPFg, Science Score, Treat, PI_SS8, BL_SS8, Latino_SS8, AI_SS8, MU_SS8, ELL_SS8, AS_SS8, ECD_SS8

g. Predictors: (Constant), MU, AI, PI, BL, AS, Latino, EconDsvntgFg, LEPPFg, Science Score, Treat, PI_SS8, BL_SS8, Latino_SS8, AI_SS8, MU_SS8, ELL_SS8, AS_SS8, ECD_SS8, ECD_Treat

h. Dependent Variable: ACT Science Score

Model	Unstandardized Coefficients		Standardized Coefficients	Sig.	95.0% Confidence Interval for B	
	B	Std. Error			Lower Bound	Upper Bound
7 (Constant)	2.365	0.363		.000	1.653	3.077
Native Am. (AI)	-0.816	3.347	-.010	.807	-7.377	5.745
Black (BL)	1.986	1.754	.046	.258	-1.452	5.424
Latino	0.137	0.733	.009	.851	-1.300	1.574
Pacific Isl. (PI)	-3.926	2.829	-.053	.165	-9.472	1.620
Asian (AS)	-1.776	0.723	-.107	.014	-3.194	-0.358
Multirace (MU)	-1.424	0.942	-.064	.131	-3.272	0.423
Econ Disadv. (ECD)	0.473	0.617	.038	.444	-0.738	1.683
Engl. Lang. Learn.	9.303	1.354	.298	.000	6.649	11.958
8 th Science Score (SS8)	1.085	0.019	.656	.000	1.048	1.122
Treat	0.742	0.106	.064	.000	0.534	0.949
AI_SS8	0.008	0.192	.002	.966	-0.368	0.385
AS_SS8	0.140	0.037	.170	.000	0.069	0.212
BL_SS8	-0.227	0.107	-.085	.035	-0.437	-0.016
Latino_SS8	-0.084	0.042	-.094	.048	-0.167	-0.001
PI_SS8	0.193	0.172	.043	.261	-0.143	0.529
MU_SS8	0.097	0.050	.081	.053	-0.001	0.195
ECD_SS8	-0.084	0.034	-.113	.014	-0.151	-0.017
ELL_SS8	-0.804	0.096	-.357	.000	-0.992	-0.616
ECD_Treat	-0.455	0.188	-.029	.016	-0.824	-0.085

APPENDIX C LOGISTIC REGRESSION MODELS

Science College Readiness

	B	S.E.	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Lower	Upper
Native American	-0.655	0.422	2.405	0.121	0.520	0.227	1.189
Black	-0.985	0.265	13.791	0.000	0.373	0.222	0.628
Hispanic/Latino	-0.723	0.095	58.284	0.000	0.485	0.403	0.584
Pacific Islander	-0.303	0.398	0.577	0.447	0.739	0.339	1.613
Asian (AS)	0.117	0.150	0.613	0.434	1.124	0.839	1.507
Multiracial	0.090	0.107	0.707	0.400	1.094	0.887	1.350
Econ Disadv.	-0.717	0.073	97.346	0.000	0.488	0.423	0.563
Engl. Lang. Learn.	-1.719	0.410	17.562	0.000	0.179	0.080	0.400
College Ready	2.354	0.064	1341.748	0.000	10.526	9.280	11.939
Treat	0.247	0.058	18.046	0.000	1.280	1.142	1.434
AS*College Ready	0.529	0.182	8.506	0.004	1.698	1.190	2.424
Constant	-1.331	0.064	427.990	0.000	0.264		

Note: R Square = .340 (Cox & Snell), .454 (Nagelkerke); Model chi square = 3194.796, $p < .001$.

STEM Interest

	B	S.E.	Wald	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Lower	Upper
Native American	-0.254	0.443	0.328	0.567	0.776	0.325	1.850
Black	-0.354	0.221	2.570	0.109	0.702	0.456	1.082
Hispanic/Latino	0.095	0.116	0.671	0.413	1.100	0.876	1.381
Pacific Islander	1.238	0.520	5.664	0.017	3.449	1.244	9.562
Asian	0.481	0.081	35.178	0.000	1.618	1.380	1.898
Multiracial	0.201	0.110	3.330	0.068	1.222	0.985	1.516
Econ Disadv.	-0.057	0.071	0.648	0.421	0.944	0.821	1.086
Engl. Lang. Learn	-0.046	0.176	0.068	0.794	0.955	0.677	1.348
STEMInt8	1.367	0.063	476.093	0.000	3.922	3.469	4.434
Treat	0.317	0.056	32.222	0.000	1.374	1.231	1.533
Latino* STEMInt8	-0.500	0.142	12.359	0.000	0.606	0.459	0.801
PI* STEMInt8	-1.495	0.752	3.949	0.047	0.224	0.051	0.980
Constant	-0.942	0.059	254.106	0.000	0.390		

Note: R Square = .111 (Cox & Snell), .148 (Nagelkerke); Model chi square = 689.975, $p < .001$.